(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 April 2004 (22.04.2004)

PCT

(10) International Publication Number
WO 2004/033641 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number:
PCT/US2003/031620

(22) International Filing Date: 3 October 2003 (03.10.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/416,002 4 October 2002 (04.10.2002) US
60/416,003 4 October 2002 (04.10.2002) US

(71) Applicant *(for all designated States except US)*: **SLOAN KETTERING INSTITUTE FOR CANCER RESEARCH** [US/US]; 1275 York Avenue, New York, NY 10021 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **SANCHEZ-CARBAYO, Marta** [US/US]; 1275 York Avenue, New-York, NY 10021 (US). **CHILDS, Geoffrey** [US/US]; 1275 York Avenue, New York, NY 10021 (US). **CORDON-CARDO, Carlos** [US/US]; 860 United Nations Plaza, 14F, New York, NY 10017 (US).

(74) Agent: **ADLER, Benjamin, A.**; Adler & Associates, 8011 Candle Lane, Houston, TX 77071 (US).

(81) Designated States *(national)*: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2004/033641 A2

(54) Title: GENE EXPRESSION PROFILING OF BLADDER CANCER

(57) Abstract: Gene expression profiling disclosed herein classified bladder tumors based on their histopathogenesis and clinical outcome. Both clusters and individual target genes were identified that would provide novel means of molecular diagnosis and outcome prediction for patients with bladder cancer.

Rec'd PCT/PTO 04 APR 2005
PCT/US2003/031620
10/530239

GENE EXPRESSION PROFILING OF BLADDER CANCER

Cross-reference to Related Applications

This non-provisional patent application claims benefit of priority of provisional patent applications 60/416,002 and 60/416,003, both filed October 4, 2002 and now abandoned.

Federal Funding Legend

This invention was supported in part by National Cancer Institute Grant CA-47538 and National Institutes of Health Grant NS39662. Consequently, the federal government has certain rights in this invention.

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention relates generally to the field of cancer research. More specifically, the present invention provides gene expression profiling for bladder cancer.

Description of the Related Art

Bladder cancer is one of the most common malignancies in developed countries, ranking as the sixth most frequent neoplasm. Bilharzial-related bladder carcinoma (BBC) is the most common malignant neoplasm in Egypt, occurring also with a high incidence in other regions of the Middle East and East Africa. Certain clinical and pathological features of bilharzial-related bladder carcinoma are different than those described for the conventional transitional cell carcinoma, such as the high incidence of detecting squamous metaplasia and the development of squamous cell carcinoma. Transitional cell carcinoma has been classified into two groups with distinct behavior and different molecular profiles: low grade tumors (always papillary

and usually superficial), and high-grade tumors (either papillary or non-papillary, and often invasive). The inactivation of both RB and p53 pathways has been shown to be required for the transformation and immortalization of uroepithelial cells, and their alterations are common and of predictive nature in clinical studies of bladder cancer. Cross-talk between these pathways and adhesion signaling, such as those generated by cadherin-catenin complexes, have been described to be involved in bladder cancer progression.

In the post genome era, and in view of the advent of high-throughput methods of molecular analysis, it is expected that specific tumor types will have distinct gene expression profiles. The elucidation of the molecular events involved in tumorigenesis and tumor progression is directly leading to the discovery and application of novel biological markers. The diagnosis and prognosis of certain neoplasms are in many cases enhanced by the use of such markers, and the marker itself may constitute a therapeutic target.

There is a need in the art for methods of gene expression profiling for bladder cancer and uses thereof. The present invention fulfills this long-standing need and desire in the art.

SUMMARY OF THE INVENTION

In the present invention, bladder cancer was characterized and new targets involved in bladder tumor progression were validated using a combination of cDNA and tissue microarray technologies. This study was designed to characterize the expression profiles of nine bladder cancer cell lines (T24, J82, 5637, HT-1376, RT4, SCaBER, TCCSUP, UMUC-3, and HT1197) using cDNA microarrays (8976 genes and ESTs). Novel targets of potential clinical relevance involved in bladder cancer progression were validated by immunohistochemistry using tissue microarrays of primary bladder tumors. Hierarchical clustering classified uroepithelial cells based on their histopathogenesis and cell cycle alterations. Keratin 10 and caveolin-1 transcripts were more abundant in tumor cells from squamous and invasive origin. Their combined expression was shown to stratify bladder tumors and define squamous differentiation.

To assess the robustness of the clustering analysis, a bootstrap resampling technique was used. This grouped tumor cell lines based on their biological properties, including cell cycle and cell adhesion features. E-cadherin, zyxin, and moesin were identified as genes differentially expressed in these clusters and related to the p53, RB and INK4A status of the cell lines. Loss of these adhesion molecules was associated with stage and grade in primary tumors, and moesin expression was also associated with survival. Deregulation of cell cycle and apoptotic pathways, such as mutations or altered expression of p53, pRB and INK4A (p16), are necessary for uroepithelial transformation. However, it appears that deregulation of cell adhesion is a common event associated with tumor progression in uroepithelial neoplasms.

Two main clusters segregating superficial from invasive transitional carcinomas were also identified that could provide prognostic information. Cytokeratin 20, neuropilin 2, p21 and p33ING1 were selected among the top ranked molecular targets differentially expressed between superficial and invasive tumors and validated by immunohistochemistry with tissue microarrays. Their expression patterns were associated with pathological stage, tumor grade, and altered RB expression. Moreover, p33ING1 expression levels were related to overall survival. Generation of a support vector machine algorithm revealed the relevance of WNT signaling and mitotic checkpoint alteration during bladder cancer progression. In summary, gene profiling successfully classified bladder tumors based on their histopathogenesis and clinical outcome, and identified molecular biomarkers of potential clinical significance.

Other and further aspects, features, and advantages of the present invention will be apparent from the following description of the presently preferred embodiments of the invention. These embodiments are given for the purpose of disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the hierarchical clustering of bladder cancer cell lines by overexpressed genes displaying red to green ratios higher than 2.0 in at least one

experiment obtained by standard clustering analysis. Using this approach, tumor cell lines were classified according to histopathogenesis criteria. SCaBER cells derived from a squamous carcinoma of the bladder were separated from cells derived from conventional transitional carcinomas. Tumor cell lines derived from invasive tumor samples were separated from cells obtained from metastatic (TCCSUP) and superficial (RT4) bladder tumors.

Figure 2 shows a representative immunostaining patterns of caveolin-1 and keratin 10 in normal urothelium, squamous metaplasia, transitional and squamous bladder tumors. Caveolin-1 expression was undetectable in normal urothelium (A), and most transitional carcinomas of the bladder (E); however, caveolin-1 levels were detected in areas of squamous metaplasia (C), as well as transitional carcinomas with squamous differentiation and squamous cell carcinomas of the bladder (G). Similarly, keratin 10 expression was undetected in normal urothelium (B), and most transitional carcinomas (F); however, keratin 10 levels were identified in squamous metaplasia (D), as well as in transitional carcinomas with squamous differentiation and squamous cell carcinomas of the bladder (H). Note the different patterns of caveolin-1 and keratin 10. Caveolin-1 was expressed in basal cell layers in the squamous metaplasia and areas of squamous carcinoma, whereas keratin 10 was identified in suprabasal layers. There was a significant difference regarding the expression of these proteins in tumors with squamous differentiation versus those with pure transitional cell carcinoma features ($p<0.001$). (Original magnifications: A through H, x200)

Figure 3 shows the robustness of the clustering analysis, a bootstrap resampling technique was applied. First, a large number (1000 in this analysis) of copies of the data were generated using a Monte Carlo resampling technique. Each of these generated datasets was then clustered using the standard hierarchical method, namely dot product (angle) metric and ward linkage. The 1000 resulting trees were then used to build a consensus tree using the CONSENS program from the PHYLIP package. The output consists of a count at each node of the tree that represents how many of the 1000 trees had that bipartition. Nodes with values close to 1000 are more significant than others scoring lower values. The higher the number at each node of the tree, the more similar the expression patterns of the cells within clusters are.

Tumor cell lines displaying similar alterations in the TP53, RB and ARF/p16 pathways grouped together within significant clusters. In the bottom part, it is shown the logarithmic ratio of the expression of these three genes in each of the cell lines. Positive and negative numbers mean higher and lower expression of these genes respectively among the cells.

Figure 4 shows a representative immunostaining patterns of zyxin, moesin and E-cadherin in primary bladder tumors. Superficial tumors showed high levels of E-cadherin (A), zyxin (B), and moesin (C) expression. However, invasive bladder neoplasms were found to express low to undetectable levels of these proteins [E-cadherin (D), zyxin (E), and moesin (F)]. There was a significant difference regarding the expression of these proteins with histopathological stage and tumor grade ($p<0.005$). (Original magnifications: A, B, and C, x200; D, E, and F, x400).

Figure 5 shows the survival analysis for patients with bladder tumors stratified by moesin expression treated as a categorical variable. The Kaplan-Meier method was used to estimate overall disease-free survival; log-rank analysis was utilized to compare the curves. Detection of moesin in the membrane of tumor cells from primary tumor samples was found to be significantly associated with overall survival in this subset of 67 bladder cancer patients (median follow up time: 36 months) ($p=0.01$).

Figure 6A shows the hierarchical clustering using bootstrap resampling techniques classified bladder tumors according to histopathological criteria. A tree is constructed by finding for each node the pairing that occurred most often in the 1000 separate trials displaying this count at each node of the tree. The number on each node represents how many times that samples to the right are grouped together out of a total of 1000 tries, a larger number indicates tight clustering. S: superficial bladder tumors, I: organ-confined invasive bladder tumors, I/M: invasive bladder tumors developing metastatic disease.

Figure 6B shows the Kaplan-Meier survival analysis of patients with superficial (cluster 1) and invasive (cluster 2) bladder tumors stratified by bootstrap clusters. Clusters containing the superficial and invasive tumors were significantly associated with overall survival (Log Rank $p=0.0025$).

Figure 6C shows a multidimensional analysis: Four groups of

expression profiles were identified by factor analysis and these were consistent with the superficial (groups 1 and 3) and invasive (groups 2 and 4) clusters.

Figure 6D shows the multidimensional analysis: the expression profiles of certain superficial tumors (163, 165, 169) were more similar to some organ-confined invasive lesions.

Figures 7A-B shows epresentative examples of the staining evaluation of p33ING1 between superficial and invasive bladder tumors. P33ING1 nuclear expression was high in superficial transitional carcinomas of the bladder (A). However, p33ING1 expression levels were lower in invasive bladder tumors (B). There was a significant difference regarding the expression of this protein in tumors regarding stage and grade in the subset of bladder cancer patients analyzed ($p<0.0005$). (Original magnifications: x400).

Figure 7C shows the Kaplan-Meier survival analysis of patients with bladder tumors stratified by the expression of p33ING1, one of the biomarkers identified in the study. p33ING1 was found to be significantly associated with overall survival in the subset of 69 bladder tumors (median follow-up time: 36 months) ($p=0.02$)

Figure 8 shows a cluster analysis based on the generated Support Vector Machine algorithm. WNT and mitotic spindle checkpoint were revealed among the altered pathways during bladder cancer progression. Relevant genes related to several networks are highlighted: p53/apoptosis (blue), kinetochore/spindle checkpoint signaling/anaphase promoting complex/cell cycle (magenta), actin polymerization/cell polarity/cell migration (orange), and vesicle budding/vesicular transport/cell adhesion (green).

DETAILED DESCRIPTION OF THE INVENTION

Large-scale transcript profiling of individual bladder tumors using cDNA array analysis disclosed herein contributed to a biologically oriented classification of bladder cancer. The combination of cDNA and tissue microarrays

facilitated validation of known and novel targets of potential clinical significance. Both clusters and individual targets provided novel means of molecular diagnosis and outcome prediction for patients with bladder cancer. Overall, gene profiling classified bladder tumors based on their histopathogenesis and clinical outcome.

Two major sets of experiments were conducted. Initially, the inventors used bladder cancer cell lines and cDNA microarrays to identify differentially expressed genes between distinct histopathological tumor types and stages of the disease. In a second approach, tissue microarrays were used to validate the potential clinical significance of the targets identified by cDNA microarrays at the microanatomical detail using immunohistochemistry on clinical material. A cohort of superficial and invasive bladder neoplasms was used to evaluate the association between molecular targets and histopathological variables including stage and grade. An additional tissue microarray, containing bladder tumors with characterized p53 and pRB alterations and annotated follow-up, was used to delineate associations between molecular markers and these critical pathways, as well as with clinical outcome.

In summary, the present invention identifies clusters and/or individual target genes that would provide novel means of molecular diagnosis and outcome prediction for patients with bladder cancer. The identified genes may also be targets for therapeutic intervention in treating bladder cancer.

In the present invention, there is provided a method of diagnosis for squamous metaplasia of bladder cancer based on the expression of caveolin-1 or keratin 10. As presented below, caveolin-1 and keratin 10 are markers of squamous differentiation. Normal urothelium and superficial conventional transitional cell carcinoma (cTCC) had undetectable levels of both caveolin-1 and keratin 10. However, areas of squamous metaplasia and carcinoma identified in bilharzial-releted bladder carcinoma, as well as areas of squamous differentiation identified in cTCC, had significant expression of both proteins.

In another aspect of the present invention, there is provided a method of diagnosis for bladder cancer based on the expression level of zyxin, E-cadherin, moesin, cytokeratin 20, neuropilin 2, p21 or p33ING1. As shown below, the expression levels of these proteins correlate with the stage and grade of bladder cancer

in individual patient.

In yet another aspect of the present invention, there is provided method of discriminating between superficial and invasive bladder cancer in an individual based on differential expression of zyxin, E-cadherin, moesin or p33ING1. Data presented below indicate that superficial tumors express high levels of zyxin, E-cadherin, moesin or p33ING1, whereas invasive neoplasms express low to undetectable levels of these proteins.

The present invention also provides a method of predicting survival outcome for a bladder cancer patient based on the level of expression of p33ING1. The level of p33ING1 expression was shown to correlate with survival outcome of bladder cancer, i.e. patients exhibiting a higher expression of p33ING1 showed a shorter survival time than those with low expression of this protein.

The present invention is also directed to a method of discriminating between superficial and invasive bladder cancer in an individual, comprising the steps of: collecting biological samples from said individual; and determining in said samples the expression of a gene identified by an accession number selected from the group consisting of AA011414, AA021434, AA021464, AA028884, AA034115, AA035095, AA043806, AA074666, AA083385, AA101348, AA127058, AA132065, AA143509, AA147928, AA156863, AA165403, AA172210, AA190401, AA256462, AA279188, AA394148, AA402766, AA421518, AA424578, AA425861, AA430520, AA434068, AA446453, AA447696, AA449831, AA450227, AA450265, AA454566, AA454862, AA455150, AA455281, AA456136, AA457092, AA457162, AA457725, AA458661, AA459663, AA464152, AA464192, AA465031, AA465378, AA465593, AA478268, AA485052, AA486313, AA486374, AA486761, AA487020, AA487223, AA487265, AA487899, AA489400, AA490047, AA490390, AA496359, AA496784, AA496948, AA504128, AA504617, AA598759, AA598815, AA620479, AA625981, AA629584, AA633757, AA669341, AA680322, AA682613, AA683085, AA705886, AA775415, AA862434, AA934762, AA935560, AI017703, H05769, H17158, H20652, H21040, H23366, H23880, H54093, H73731, H84444, H93463, H94897, H99502, N30811, N54338, N69283, N73536, N91962, R16165, R22439, R24543, R25377, R27552, R49144,

R53889, R55763, R69307, R76314, R78514, T53404, T57815, T67053, T81091, T96829, W49619, W69906, W96107, AA010393, AA019591, AA024832, AA113339, AA115248, AA121704, AA134595, AA142875, AA157797, AA165400, AA284268, AA284292, AA404694, AA406603, AA421783, AA424834, AA429399, AA431184, AA435936, AA436158, AA436871, AA443193, AA443285, AA453607, AA453748, AA454579, AA454625, AA455119, AA457374, AA459950, AA460365, AA463958, AA482325, AA488526, AA488645, AA489246, AA489661, AA496780, AA504894, AA599093, AA609067, AA609134, AA621335, AA705060, AA708310, H09747, H09818, H10335, H17335, H23277, H29292, H41096, H53141, H58736, H65834, H70815, H93463, H95989, N21548, N38891, N56882, N58283, N66933, N94428, R08891, R09585, R22271, R28669, R36449, R43525, R44132, R51080, R56219, R56432, R60053, R60927, R64066, R92455, R94943, R98628, R99918, T50370, T55592, T61792, T68461, T71680, T86983, T90641, T96711, W31919, W56308, W99364, wherein said gene is differentially expressed at the mRNA level in superficial and invasive bladder cancer.

The present invention if further directed to a method for identifying the presence or absence of a squamous metaplasia of bladder cancer phenotype in a cell or cells, comprising determining the expression level of caveolin-1 or keratin 10 in said cell or cells, wherein a detectable expression level of caveolin-1 or keratin 10 in the cell or cells indicates the presence of squamous metaplasia of bladder cancer phenotype and an undetectable level of caveolin-1 or keratin 10 in said cell or cells indicates the absence of squamous metaplasia of bladder cancer phenotype.

In another aspect, the present invention provides a method of identifying the presence or absence of a squamous metaplasia of bladder cancer in an individual, comprising the steps of: collecting a biological sample from said individual; and determining the expression level of caveolin-1 or keratin 10 in said sample, wherein a detectable expression level of caveolin-1 or keratin 10 in said sample indicates the presence of said squamous metaplasia of bladder cancer and an undetectable level of caveolin-1 or keratin 10 in said sample indicates the absence of said squamous metaplasia of bladder cancer. Preferably, the expression level is a protein expression level or a nucleic acid expression level.

In another aspect, the present invention provides a method of identifying the presence or absence of a bladder cancer in an individual, comprising the steps of: collecting a biological sample from said individual; and determining in said sample the level of expression of a protein selected from the group consisting of zyxin, E-cadherin, moesin, cytokeratin 20, neuropilin 2, p21 and p33ING1, wherein the level of expression indicates the presence or absence of a bladder cancer in the individual. In another specific aspect, this method may further comprise correlating the expression level with the stage and grade of bladder cancer in the individual. Preferably, the expression of the protein is determined at protein level or the expression of the protein is determined at nucleic acid level.

In another aspect, the present invention provides a method of discriminating between a superficial and an invasive bladder cancer in an individual, comprising the steps of: collecting a biological sample from said individual; and determining in said sample the level of expression of a protein selected from the group consisting of zyxin, E-cadherin, moesin and p33ING1, wherein said protein is differentially expressed in superficial and invasive bladder cancer. Preferably, the level of expression is a level of expressed protein or the level of expression is a level of nucleic acid expression.

In another aspect, the present invention provides a method of predicting survival outcome of an individual having bladder cancer, comprising the steps of: collecting biological samples from said individual; and determining in said samples the level of expression of p33ING1, wherein said level of expression correlates with survival outcome of said individual. Preferably, the expression of said protein is determined at protein or nucleic acid level.

In another aspect, the present invention provides a kit for identifying the presence or absence of a squamous metaplasia of bladder cancer phenotype in a cell or cells, comprising a reagent or reagents capable of determining the expression level of caveolin-1 or keratin 10 in said cell or cells, wherein detectable expression levels of caveolin-1 or keratin 10 in said samples indicates the presence of squamous metaplasia of bladder cancer phenotype and undetectable levels of caveolin-1 or keratin 10 in said cell or cells indicates the absence of squamous metaplasia of bladder cancer phenotype.

In another aspect, the present invention provides a kit for identifying the presence or absence of bladder cancer in an individual, the kit comprising a reagent or reagents capable of determining the level of expression of a protein selected from the group consisting of zyxin, E-cadherin, moesin, cytokeratin 20, neuropilin 2, p21 and p33ING1.

In another aspect, the present invention provides a kit for discriminating between superficial and invasive bladder cancer in an individual, the kit comprising a reagent or reagents capable of determining the level of expression of a protein selected from the group consisting of zyxin, E-cadherin, moesin and p33ING1.

In another aspect, the present invention provides a kit for predicting survival outcome of an individual having bladder cancer, the kit comprising a reagent or reagents capable of determining in the samples the level of expression of p33ING1.

In all embodiments of the kit of the present invention, it is contemplated that the reagent could be an antibody. Alternatively, in all embodiments of the kit of the present invention, it is contemplated that the one reagent could be a nucleic acid. In the embodiments of the kit of the present invention, it is contemplated that the it may further compriseinstructions for correlating the level of expression with a clinical diagnosis.

The following examples are given for the purpose of illustrating various embodiments of the invention and are not meant to limit the present invention in any fashion. The present examples, along with the methods, procedures, treatments, molecules, and specific compounds described herein are presently representative of preferred embodiments. One skilled in the art will appreciate readily that the present invention is well adapted to carry out the objects and obtain the ends and advantages mentioned, as well as those objects, ends and advantages inherent herein. Changes therein and other uses which are encompassed within the spirit of the invention as defined by the scope of the claims will occur to those skilled in the art.

EXAMPLE 1

Molecular Profiling of Bladder Cancer Using cDNA Microarrays Defines

Histogenesis And Biological Phenotypes

Expression profiling classified nine bladder cancer cell lines under study based on histopathological characteristics of the tumors from which they were obtained. Keratin 10 and caveolin-1 expression was associated with the presence of squamous differentiation, as well as with pathological stage and tumor grade. The application of bootstrapping techniques to hierarchical clustering grouped most of the bladder cancer cell lines based on their alterations in p53 and RB pathways. Target genes identified herein, namely zyxin, E-cadherin and moesin, were associated with p53 and/or pRB alterations in the primary bladder tumors analyzed. Identified target genes obtained from high-throughput molecular profiling of cultured cells were shown to have clinical impact when validated in primary bladder tumors using tissue microarrays. These target genes were significantly associated with bladder cancer progression; and moesin provided predictive outcome information.

Cell Culture And RNA Extraction

Nine bladder cancer cell lines including T24, J82, 5637, HT1376, RT4, SCaBER, TCCSUP, UMUC-3, and HT1197 were obtained from ATCC (Rockville, MD) and cultured under identical conditions following standard procedures. All cells were grown and harvested at 75%-90% confluence no longer than 4-6 passages in culture for the extraction of total RNA using RNeasy protocol (Qiagen, Valencia, CA). Cytospins were also prepared and later used for target validation.

Preparation of cDNA Microarrays And Image Acquisition

A set of 8976 sequence-verified human IMAGE cDNA clones, representing both known genes and expressed sequence tags (ESTs), were PCR amplified and spotted onto poly-lysine coated microscope slides using a custom robot designed and built at Albert Einstein College of Medicine (Cheung et al., 1999).

Ten μg of total RNA of each cell line was labeled with Cy5 (red) and hybridized against 10μg of total RNA of a pool containing equal RNA quantities of all of these cell lines labeled with Cy3 (green). Labeling and hybridization of cDNA to arrays was carried out as previously described (Stears et al., 2000). One duplicate or one reverse-labeling experiment was carried out for validation of expression changes

of the hybridization of the cell lines. Following hybridization, slides were washed, dried and scanned by a custom-built laser scanner (Cheung et al., 1999). Intensity data were integrated with 8x oversampling (Cheung et al., 1999). Scanalyse software was used for gridding and calculation of red (R) and green (G) signal intensities (Eisen et al., 1998).

Collection And Analysis of The Data of The cDNA Microarrays

Normalization: Before any analysis, plots of the fold change versus the average intensity were examined to look for abnormalities in single-array data. It is common to plot a red versus green channel scatter plot to examine distribution of intensities; however, the inventors found that transforming to fold change versus average intensity displayed the data in a more easily viewed form. If Ired is the background subtracted red channel intensity, and Igreen is the background subtracted green intensity, then the following variables were created: R=Ired/Igreen and A=√(Ired x Igreen), where R is simply the fold change ratio and A is the average intensity (the geometric mean which is equivalent to averaging the log intensity). The curvature in the scatter plot indicated a dependence of the ratio R on the overall intensity. This curve was then used to normalize the data: logIred/Igreen->log (Ired/Igreen)-c(A) where c(A) is the fit. This is equivalent to multiplying the green channel intensity (or dividing the red) by an intensity dependent normalization constant k(A) where log[(k(A)]= c(A). Optimal normalized data should be horizontal and centered at zero. Samples were normalized using this intensity-dependent normalization using the Splus function lowess (Dudoit, 2000). Normalized fold changes in gene expression were then used to further analyze and cluster the various cell lines.

Cutoffs: The absolute value of the fold change (R/G and G/R) had to be greater than 2.0 in at least one experiment, and the average intensity (A) need to be greater than 300. This filter reduced the number of genes from 8976 to 234. Data was filtered to select genes that had both a fold change to remove the background of mostly unchanging genes and an average intensity distinguishable from the noise of the microchip hybridization.

Clustering: The relationship among cell lines was analyzed using hierarchical clustering limiting to over-expressed genes with R/G or G/R ratios higher

than 2.0 in at least one experiment (Eisen et al., 1998). To assess the robustness of the clustering analysis, a bootstrap resampling technique was used to generate 1000 copies of the data set by adding Gaussian noise to the original data. The mean value of the noise was zero and the standard deviation was dependent on the average intensity of a given spot. In order to determine this intensity dependent noise value, the data from a sample replicated eight times were used. By fitting a curve to the scatter of the standard deviation of the eight replicates as a function of average log intensity, the inventors obtained a curve for the average noise as a function of intensity. This was used in the Montecarlo resampling to set the value of the standard deviation. Each of the 1000 bootstrap samples was then clustered using the hierarchical method with the dot plot product (angle) metric and ward linkage. A consensus tree was constructed using the CONSENS program (version 3.5c). This program constructs a tree by finding for each node the pairing that occurred most often in the 1000 separate trials. A graph was constructed that displays this count at each node of the tree. Nodes with values closer to 1000 are more robust than with lower ones.

Validation of The Results

Northern Blotting: Northern hybridization was performed using 10 ug of total RNA from the bladder cancer cell lines used in the analysis (see above), and probes generated from the cDNA clones (data not shown).

Tissue Samples And Tissue Microarrays: Three different bladder cancer microarrays were constructed for this study. Normal and tumor tissues were embedded in paraffin and five-μm sections were stained with hematoxylin and eosin to identify viable, morphologically representative areas of the specimen from which needle core samples were taken using a precision instrument (Beecher Instruments, Silver Spring, MD) (Hoos et al., 2001). From each specimen triplicate tissue cores with diameters of 0.6 mm were punched and arrayed on the recipient paraffin block. Five-μm sections of these tissue array blocks were cut and placed on charged polylysine-coated slides and used for immunohistochemical analysis. Arrayed normal tissues known to express the antigens under study served were used as baseline positive controls and showed physiological expression patterns of these markers.

These tissue microarrays included a total of 173 bladder primary transitional cell carcinoma tumors obtained under Institutional Review Board approved protocols. Tumor stage and grade were defined according to consensus criteria. A total of 40 superficial and 64 invasive transitional cell carcinoma tumors were analyzed in two microarrays. These tumors corresponded to 14 grade 1, 8 grade 2, and 82 grade 3 lesions. Another tissue microarray comprised a cohort of 69 bladder primary transitional cell carcinoma cases with known p53, p16 and pRB status, and consisted of two superficial and 67 invasive lesions. In addition, 20 cases of Bilharzial-related invasive bladder cancer (BBC) were also analyzed, including 14 squamous (S-BBC) and 6 transitional (T-BBC) carcinomas, for a total of 193 cases. These BBC lesions were also analyzed for patterns of p53 expression (see below).

Immunohistochemistry: Protein patterns of expression of identified targets were assessed at the microanatomical level for caveolin-1, keratin 10, E-cadherin, zyxin, and moesin, using both cytospin from all cell lines studied and tissue samples outlined above. Standard immunoperoxidase procedures were used for immunohistochemistry (Hoos et al., 2001). The following antibodies were used: anti-caveolin-1, mouse monoclonal IgG1 at 1:1000 dilution (2.5 μg/ml) (BD Transductions Labs, Lexington, KY); anti-keratin 10, mouse monoclonal clone DC-K10 at 1:2000 (1.0 μg/ml) (Neomarkers, Fremont, CA); anti-E-cadherin, mouse monoclonal clone 36 at 1:1000 (2.5 μg/ml) (BD Transductions Labs, Lexington, KY); anti-moesin, mouse monoclonal clone 38/87 at 1:50 (4 μg/ml) with microwave pretreatment of the slides (Neomarkers, Fremont, CA); anti-zyxin, mouse monoclonal clone 21 at 1:25 (10 μg/ml) with microwave pretreatment of the slides (Transduction Labs, Lexington, KY); anti-RB, mouse monoclonal clone 3C8 at a final concentration of 1.2 μg/ml (QED Bioscience, San Diego, CA); anti-p16, mouse monoclonal clone clone DCS-50.1/H4 at 2.5 μg/ml (Calbiochem, Cambridge, MA); and a mouse anti-human monoclonal antibody against p53 (1:500, Ab-2, clone 1801; Calbiochem). Staining conditions were optimized on sections from formalin-fixed, paraffin-embedded tissue controls for each antibody as specified by manufacturers. Antibody reactivity was detected using diaminobenzidine as chromogen, and sections were counterstained with hematoxylin. The primary antibody was omitted for negative controls. p53 staining

was defined as negative (undetectable levels to ≤20 of tumor cells displaying nuclear staining) or positive (moderate to intense nuclear immunoreactivities in >20% of cells) (McShane et al., 2000). There is no consensus on the cutoffs of the immunohistochemical expression of the other markers, and thus they were analyzed as continuous variables, or taking the cutoff of 0% versus higher than 0% when they were considered as categorical.

Data analysis: All conventional transitional cell carcinoma (n = 173) were used for the analysis of association between p53 and pRB with keratin 10, caveolin-1, E-cadherin, zyxin and moesin. These cases were also utilized for evaluating marker expression versus histopathological stage and tumor grade using the non-parametric Wilcoxon-Mann-Whitney and Kruskall-Wallis tests (Tudor and Koch, 1994). The consensus value of the three representative cores from each tumor sample arrayed was used for statistical analyses. The association of keratin 10 and caveolin-1 with squamous differentiation was analyzed using the total cohort of 193 cases, including the 20 Bilharzial-related bladder tumors. Expression values were displayed as mean values accompanied of 95% confidence intervals and/or range.

The relationship of marker to outcome was evaluated using a subset of 69 conventional transitional cell carcinoma cases for which follow up was available. Overall-survival time was defined as the months elapsed between transurethral resection (two superficial lesions) or cystectomy (rest of cases) and death from disease (or the last follow-up date). Patients who were alive at the last follow-up or lost to follow-up were censored. For survival analysis, expression marker results were analyzed as continuous variables. Membrane expression of moesin was also considered as a categorical variable becasue its median expression value was zero. The association of the marker expression levels with overall survival was analyzed using the Wald test, and the log-rank test was used to examine their relationship when different cutoffs were applied (Fleming and Lin, 2000). Survival curves were plotted using standard Kaplan-Meier methodology (Kaplan and Meier, 1958). Additionally, the association of the markers with the p53 (mutation analysis) and p16 (mutation and polymorphism analysis) status in this subset of 69 patients was evaluated. Associations between markers were analyzed using Kendall's tau test.

Histopathogenetic Categorization of Bladder Cancer Cell Lines

Hierarchical clustering of cDNA microarray experiments, based on 234 genes that showed a R/G or G/R fold ratio higher than 2 and intensities higher than 300 classified these tumor cells according to the histopathological characteristics from the tumors they were obtained from. SCaBER cells, derived from a squamous carcinoma of the bladder, were distinguished from cells derived from transitional carcinomas. Moreover, tumor cells from invasive lesions clustered together, and were separated from those cells derived from metastatic (TCCSUP) or superficial (RT4) bladder cancers (Figure 1). The complete list of 234 genes is shown in Table 1.

Caveolin-1 And Keratin 10 Are Markers of Squamous Differentiation And Are Associated With Tumor Stage And Grade

Both caveolin-1 and keratin 10 were differentially expressed among the various bladder tumor cells lines analyzed. High levels of caveolin-1 were detected in SCaBER cells, while its expression was low to undetectable in RT4 cells. The expression of Keratin 10 was high in several cell lines, and a previous report had linked its expression to squamous bladder carcinoma.

Based on these observations and availability of well-characterized antibodies to their encoded products, the inventors further explored their patterns of expression in several normal samples of human urothelium, urothelial squamous metaplasia, and the above-mentioned tissue collections, including 173 conventional transitional cell carcinoma (cTCC) and 20 bilharzial-related bladder carcinoma (BBC) (16 squamous-BBC and 4 transitional-BBC). It was found that normal urothelium and superficial cTCC had undetectable levels of both caveolin-1 and keratin 10. However, areas of squamous metaplasia and carcinoma identified in BBC, as well as areas of squamous differentiation identified in cTCC, had significant expression of both proteins. Caveolin-1 was expressed in 12 of 16 of squamous-BBC, as well as in 2 of 4 of transitional-BBC and 72 of 173 of cTCC. Keratin 10 was found in 8 of 16 of squamous-BBC, 1 of 4 of transitional-BBC, and 28 of 173 of cTCC (Figure 2). Statistical analysis of data revealed that both caveolin-1 and keratin 10 were significantly associated with identification of squamous differentiation in 49 of 193

patients ($p<0.001$).

Separate analysis of the 173 cTCC lesions revealed that only one of 42 superficial lesions displayed caveolin-1 in few tumor cells (approximately 3% tumor cells), while 70 of 131 invasive tumors showed caveolin-1 expression ranging from 3% to 83% tumor cells. Keratin 10 was undetectable in all superficial lesions, while 28 of 131 invasive tumors displayed immunoreactivities ranging from 3% to 70% tumor cells. All grade 1 lesions showed undetectable expression of caveolin-1 and keratin 10. Grade 2 tumors showed expression lower than 3% for caveolin-1, while in grade 3 tumors the mean number of cells showing positive expression of caveolin-1 was 13% (95% CI :8.9-17.2%). Keratin 10 expression was also undetectable for tumors of grade 2 and for those displaying grade 3 the mean number of cells showing positive expression of keratin 10 was 2.3% (95% CI :0.5-4.1%). Overall, there was a statistical association between caveolin-1 expression and both tumor stage ($p<0.001$) and grade ($p<0.001$). Keratin 10 also reached a significant statistical association with tumor stage ($p=0.019$) and grade ($p=0.018$).

It is postulated that cTCC in which these products were identified harbor morphologically unrecognizable areas of squamous differentiation. This could have important clinical implications since it has been reported that invasive bladder tumors with squamous features do not respond to MVAC treatment.

Clustering Associates Expression Profiling To Biological Phenotypes Related To p53 And RB Pathways.

A Montecarlo bootstrap method was applied to establish the robustness among the grouping of the cell lines based on the expression of 234 genes selected from the cDNA microarrays showing a R/G or a G/R fold ratio higher than 2 and intensities higher than 300. Using this approach, the analyzed tumor cells assembled based on their reported molecular alterations related to the p53 and RB signaling pathways (Figure 3). Two main clusters or groups were identified: T24, SCABER and UMUC3 (group 1) and HT1376, HT1197 and TCCSUP (group 2). Briefly, cells that harbor TP53 mutations at exons 4 and 5, detectable pRB, and INK4A mutations (group 1) clustered together and were distinguishable from those having TP53 mutations affecting exons 7, 10, and 11, undetectable levels of pRB, and

a wild-type INK4A locus (group 2). Cells with p53 mutations in exon 8 showed a higher distance from this second group.

In an attempt to identify genes related to this clustering and of potential biological significance, a search for common over- and under-expressed genes in each cluster-group and not in the other was performed. Of interest, only three known genes (zyxin, protocadherin 13, and moesin) and an EST were found differentially expressed between the two significant clusters. In order to validate these results at the protein level, immunohistochemical studies using antibodies to zyxin, moesin, and E-cadherin (a down stream product of protocadherin 13 for which antibodies are not available) were performed on cytospins from the analyzed bladder cancer cell lines. Cells from group 1 displayed lower transcript levels of protocadherin 13 and zyxin than cells from group 2. Alternatively, group 2 cells had lower transcript levels of moesin than cells from group 1.

Zyxin, E-cadherin, And Moesin Are Associated With Altered pRB Expression, TP53 Mutation Localization, Tumor Stage And Grade In Primary Bladder Cancer

In order to validate the results obtained through permutation clustering using clinical primary bladder tumors, patterns of p53 and pRB expression were assessed in the cohort of patients under study. In addition, TP53 mutation status was established in a subset of 69 cTCC cases: 37 tumors had wild-type TP53, 9 lesions harbored mutations affecting exons 4 or 5, and 23 had a mutation between exons 6 and 11 (Wikman et al., 2000). Cytoplasmic zyxin expression was significantly associated with detection of TP53 mutations affecting exons 6-11 ($p=0.03$). Regarding p53 and pRB expression status: 73 cases displayed a p53 positive phenotype (nuclear immunoreactivities $\geq 20\%$ tumor cells), while the remaining cases were classified as having a p53 negative phenotype; 44 cases had undetectable pRB, while the remaining cases showed heterogeneous pRB nuclear immunoreactivities. Zyxin was observed in the cytoplasm of 102 of 173 cTCC cases, while membrane E-cadherin staining was found in 146 of these 173 lesions (Figure 4). Moesin was detected as a membrane staining in 26 of 173 cTCC (Figure 4).

No statistical association was found between expression patterns of p53 and either zyxin, E-cadherin or moesin. Nor was there any relationship among

mutations, polymorphisms or protein over-expression of p16 in a subset of 69 patients of whom INK4A/p16 was available. However, there was a significant association between pRB levels and these three markers (Table 2). Furthermore, expression levels of E-cadherin, zyxin and moesin were significantly associated with tumor stage ($p<0.001$ for the three markers) and grade ($p<0.001$ for E-cadherin and zyxin, and $p=0.005$ for moesin).

Moesin Is A Predictive Marker In Bladder Cancer

The potential overall survival prognostic utility of caveolin-1, keratin 10, E-cadherin, moesin, and zyxin was evaluated using 69 cTCC for which clinical follow-up was available. Membrane moesin expression was associated with overall survival ($p=0.01$) in this subset of 69 patients (Figure 5). Ppatients presenting a positive moesin expression displayed p53≥20% in 5/10, E-cadherin>0% in 10/10, zyxin>0% in 6/10, total pRb≥10% in 9/10, underphosphorylated pRb≥10% in 3/10 and p16>0% in 2/10 of the cases. No polymorphism or mutation in INK4a/p16 gene was detected.

Bladder cancer comprises a variety of distinct neoplastic disorders. Transitional and squamous carcinomas are the most prevalent forms of bladder cancer. However, adenocarcinomas, small cell, and neuroendocrine tumors are also found as primary bladder tumors with a lower frequency. Identification of the prevailing and, if present, secondary histogenetic features of the tumor have significant clinical connotations, since it is well known the lack of response to certain therapeutic regimens in the context of specific tumor types. For example, squamous carcinoma of the bladder has been reported to be more resistant to radio- and chemotherapy than conventional transitional bladder tumors. Data from this study revealed a characteristic pattern of caveolin-1 and keratin 10 expression in early squamous metaplasia and squamous carcinomas in the setting of BBC. In addition, the inventors observed that the expression of caveolin-1 and keratin 10 in certain cTCC, usually identified as clusters of tumor cells heterogeneously stained within the bulk of the tumor. Furthermore, there was a significant association regarding detection of both caveolin-1 and keratin 10 in the bladder tumor samples analyzed. Thus, they may serve as markers of squamous differentiation prior to the morphological identification

of this cellular phenotype. This phenomenon might be linked to the lower response to MVAC observed in certain cTCC, harboring histologically unrecognizable areas of squamous differentiation, and may be of assistance in selecting those patients that would benefit from other therapeutic regimens.

The increased expression of caveolin-1 has been related to cellular transformation and tumor progression, being associated with augmented cell signaling activity. One of the molecules that is organized and concentrated in the scaffolding domain of caveolin-1 is the epidermal growth factor receptor (EGFr). Bladder cancer cells have increased growth factor receptors, including EGFr and Her-2/Neu proteins, and this phenomenon has been associated with tumor progression. Interestingly, data from the present analysis also links caveolin-1 expression with increased tumor pathological stage and tumor grade. A recent study has also described association with tumor grade but not with patient outcome. Whether the role of caveolin-1 as a membrane protein implicated in selective transcytosis and increased signaling, or to what extent keratinization are critical in bladder cancer progression and chemoresistance in tumors presenting squamous differentiation has to be further studied.

One of the most notable findings presented above is that significant clusters obtained by bootstrapping methods grouped the bladder cancer cell lines analyzed based on their p53 and RB pathway status. Growth control in mammalian cells is accomplished largely by the action of pRB (regulating exit from the G1 phase) and the p53 protein (triggering growth arrest or apoptotic processes in response to cellular stress). pRB and p53 serve collaborative roles in tumorigenesis as evidenced by their frequent alterations in human tumors, including bladder cancer. The mechanistic basis for this dual requirement stems, in part, from the deactivation of a p53-dependent cell suicide program that would normally be brought about as a response to unchecked cellular proliferation resulting from RB-deficiency.

Two significant major patterns arise from the bootstrapping analysis study in these cell lines based on the combined expression of 234 genes and contrasted with the molecular characterization of p53/pRB/INK4a in these cells lines already described. The combined alteration of these critical networks could support these clusters. Cells harboring TP53 mutations in the amino-terminal transactivation

domain, presenting with high pRB levels, and INK4A mutations grouped together (UMUC3, SCaBER, and T24); and certain cells with TP53 mutations in the core domain, undetectable pRB levels, and wild-type INK4A locus were in the same cluster (HT1197, HT1376, and TCCSUP). TP53 mutations in the core domain affect the ability of p53 to bind DNA and are associated with loss of the contralateral allele, completely inactivating p53 function and thus impacting on both cell cycle arrest and induction of apoptosis. However, mutations in the transactivation domain result in products that preserve to some extent p53 activities, such as DNA binding. This incomplete p53 suppressive phenotype is usually associated with other alterations in the pathway, mainly p14ARF mutations or Hdm2 amplification/overexpression. In these cases, the inventors found detectable pRB expression, but the other prevalent mutation of the RB pathway, namely p16/INK4A deletions, was detected in all cell lines displaying such genotype. Finally, the inventors observed that RT4 cells had a wild-type TP53, but harbored a homozygous INK4A deletion, and lacked pRB expression. These data support the working model previously reported. However, cells with mutations in exon 8 in the core domain (5637 and J82) were not included into the cluster described by the boostrap technique.

Three adhesion-related molecules, zyxin, E-cadherin, and moesin, were found to be associated with the p53/RB patterns discussed above. Those lines harboring TP53 mutations in the core domain and lacking pRB displayed low moesin transcript levels; while those lines with TP53 mutations in the transactivating domain and high levels of pRB displayed low zyxin and E-cadherin. Moreover, the associations observed in cell lines were also found and validated in primary bladder tumors. Low levels of moesin, zyxin and E-cadherin were significantly associated with advanced pathological stage and higher tumor grade, supporting their involvement in bladder cancer progression.

Alterations of E-cadherin had been described as common events in bladder cancer. The present observations regarding the association of E-cadherin with histopathological stage and tumor grade were in accordance with other previous studies. Loss of zyxin has been associated with neoplastic transformation, as it was found as a marker of acute myeloid leukemia subtype. The functional role of zyxin is not completely defined, but it appears to be involved in signaling networks

established between focal adhesion plaques and the nucleus. There has been no report dealing with altered patterns of zyxin in clinical samples to date. Results from this study revealed that moesin provided prognostic information regarding poor outcome in bladder cancer. The loss of moesin has recently been reported as being associated with advanced ovarian cancer, metastatic melanoma, and lung carcinoma, suggesting that it could have a critical tumor suppressive function altered in multiple human cancers. It was also observed a predictive utility of moesin expression in bladder cancer. The inventors evaluated the alterations in p53/pRB/p16 in these patients that showed a shorter survival. Most of them presented alterations in these molecules that may also account with their aggressive outcome.

The three adhesion-related markers (zyxin, E-cadherin, moesin) found differentially expressed in bladder cancer in the present study share a common feature: a relationship to the beta-catenin pathway. Beta-Catenin is a cytoplasmic protein that participates in the assembly of cell-cell adherens junctions by binding cadherins to the actin cytoskeleton. The cytoplasmic domain of E-cadherin interacts directly with beta-catenin. Zyxin is a cytoplasmic adherens junction protein found in complexes with alpha-actinin and actin. The association of zyxin with cadherins has also been previously reported. Moesin is a member of the ERM (ezrin, radixin and moesin) family of proteins located just beneath the plasma membranes, which are also thought to be involved in the association of actin filaments with the plasma membrane regulating cell-cell and cell-matrix adhesion. The association of ezrin with E-cadherin and beta-catenin has also been revealed by coprecipitation studies. The involvement of beta catenin in bladder cancer progression has recently been described in murine and human models, an observation supporting the present findings. Among downstream targets of beta-catenin, cyclin D1 has been shown to be critical in G1-S cell cycle transition by phosphorylating pRB. Interestingly, the three identified markers were associated with RB gene expression in primary tumors, and this association could be attributed to alteration of cyclin D1 expression levels as reported in colorectal and desmoid tumors.

Disruption of the beta-catenin signaling pathway by alterations in the physiological balance between its interactions with zyxin, E-cadherin or moesin could mechanistically account for the invasiveness potential that certain bladder cancer cell

lines under study display. Other alternative mechanisms affecting this pathway include alterations of the Wnt signaling, RAS mutations, and mutations affecting the beta-catenin gene itself. Aberrant accumulation of beta-catenin in solid tumors has been also associated with mutational inactivation of the TP53 gene. Overexpression of wild-type p53, by either transfection or DNA damage, has been shown to down-regulate beta-catenin in human and mouse cells. However, whether the association between zyxin and TP53 mutations could also be related to an altered beta-catenin pathway remains to be elucidated.

The present study has identified deregulation of three adhesion molecular targets as common alterations in high-grade bladder cancer cells with different described phenotypes of cell cycle regulator genes. The loss of these cell-adhesion molecules was correlated with tumor progression in primary bladder tumors. This observation reveals the importance of the interactions among tumor cells as well as tumor cells with the surrounding stroma in cancer progression. It appears that deregulation of cell cycle and apoptotic pathways, such as mutations or altered expression of p53 and pRB, are necessary for uroepithelial transformation. However, they appear to be insufficient for bladder cancer progression. Even though linear models are a "simplification" of complex pathological events, it appears that deregulation of cell adhesion is a common event associated with tumor progression in bladder cancer, independently of the genetic alterations triggering tumorigenesis. Expression profiling has revealed the common deregulation of cell adhesion displayed in highly invasive and differentiated cells alteration of alternative adhesion pathways. The most relevant finding was that these molecular targets identified in vitro, zyxin, E-cadherin and moesin, were found critical in progression in clinical material supporting a relevant role of deregulation of cell adhesion in bladder cancer progression.

In summary, molecular profiling using cDNA microarrays clustered bladder cancer based on both histopathogenesis and biological criteria. Novel targets genes have been validated using tissue arrays containing well characterized primary tumors. Keratin 10 and caveolin-1 defined squamous differentiation, and might become useful markers to further stratify bladder tumors. E-cadherin, moesin and zyxin were associated with tumor progression, revealing the relevance of deregulation

of cell adhesion in bladder cancer progression. Finally, moesin expression appeared to be a significant prognostic factor associated with patient survival.

TABLE 1

234 Genes Used For Hierarchical Clustering Analysis of Bladder Cancer Cell Lines

cDNA Microarray

1. Non-specific cross reacting antigen
2. Nerve growth factor beta
3. Myelin basic protein
4. Membrane metallo-endopeptidase (neutral endopeptidase, enkephalinase, CALLA, CD10)
5. Human stanniocalcin precursor (STC) mRNA, complete cds
6. Human mRNA for KIAA0269 gene, complete cds
7. Human clone 23587 mRNA sequence
8. Homo sapiens mRNA for NB thymosin beta, complete cds
9. Homo sapiens Rac3 (RAC3) mRNA, complete cds
10. Fibrillin 2
11. Brain-derived neurotrophic factor
12. Plasminogen activator inhibitor, type II (arginine-serpin)
13. MULTIFUNCTIONAL AMINOACYL-TRNA SYNTHETASE
14. MHC class II DQ-beta associated with DR2, DQw1 protein
15. Keratin 5 (epidermolysis bullosa simplex, Dowling-Meara/Kobner/Weber-Cockayne types)
16. Human skeletal muscle LIM-protein SLIM1 mRNA, complete cds
17. Human 19.8 kDa protein mRNA, complete cds
18. Heme oxygenase (decycling) 1
19. ESTs, Weakly similar to !!!! ALU SUBFAMILY SX WARNING ENTRY !!!! [H.sapiens]
20. ESTs, Highly similar to IG ALPHA-2 CHAIN C REGION [H.sapiens]
21. ESTs
22. ESTs
23. ESTs
24. ESTs
25. ESTs, Weakly similar to F43C1.3 [C.elegans]
26. ESTs, Highly similar to ZYXIN [Gallus gallus]
27. ESTs, Highly similar to ZINC FINGER PROTEIN 42 [Homo sapiens]
28. ESTs, Highly similar to INHIBIN BETA A CHAIN PRECURSOR [Bos taurus]
29. ESTs
30. ESTs
31. ESTs
32. ESTs
33. ESTs
34. ESTs
35. ESTs
36. ESTs

37. ESTs
38. ESTs
39. ESTs
40. ESTs
41. ESTs
42. ESTs
43. ESTs
44. ESTs
45. ESTs
46. INTERFERON-INDUCIBLE PROTEIN 9-27
47. Human tumor susceptiblity protein (TSG101) mRNA, complete cds
48. Human mRNA for KIAA0075 gene, partial cds
49. Human mRNA for BST-2, complete cds
50. Human TRAF-interacting protein I-TRAF mRNA, complete cds
51. Homo sapiens mRNA for zinc finger protein FPM315, complete cds
52. H.sapiens mRNA for processing a-glucosidase I
53. Complement component 4-binding protein, beta
54. Casein kinase 2, alpha prime polypeptide
55. UDP-GLUCURONOSYLTRANSFERASE 2B4 PRECURSOR, MICROSOMAL
56. ESTs, Highly similar to RAS-RELATED PROTEIN RAB-18A [Lymnaea stagnalis]
57. ESTs, Highly similar to 50S RIBOSOMAL PROTEIN L2 [Bacillus stearothermophilus]
58. Human osteoprotegerin (OPG) mRNA, complete cds
59. ESTs, Weakly similar to weak similarity to ribosome releasing factors [C.elegans]
60. 5' nucleotidase (CD73)
61. S100 calcium-binding protein A9 (calgranulin B)
62. Human G protein gamma-11 subunit mRNA, complete cds
63. ESTs
64. ESTs
65. X-LINKED HELICASE II
66. Protein-tyrosine kinase 7
67. Pregnancy-specific beta-1 glycoprotein 13
68. Peripheral myelin protein 22
69. Moesin
70. Membrane component, chromosome 1, surface marker 1 (40kD glycoprotein, identified by monoclonal antibody GA733)
71. Interferon-inducible 56-KDa protein
72. INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 1 PRECURSOR
73. Human putative EPH-related PTK receptor ligand LERK-8 (Eplg8) mRNA, complete cds
74. Human clone 23589 mRNA sequence
75. Homo sapiens mRNA for aurora/IPL1-related kinase, complete cds
76. H.sapiens IL-13Ra mRNA
77. GLYCYLPEPTIDE N-TETRADECANOYLTRANSFERASE
78. Fibronectin 1
79. Annexin VIII
80. Keratin 10 (epidermolytic hyperkeratosis; keratosis palmaris et plantaris)

81. KINESIN HEAVY CHAIN
82. Human (clone 8B1) Br-cadherin mRNA, complete cds
83. Homo sapiens mRNA for osteoblast specific factor 2 (OSF-2os)
84. Hemopoietic cell kinase
85. H.sapiens mRNA for testican
86. ESTs, Highly similar to METALLOTHIONEIN-II [H.sapiens]
87. ESTs, Highly similar to DEVELOPMENTAL PROTEIN SEVEN IN ABSENTIA [Drosophila melanogaster]
88. ESTs
89. ESTs, Moderately similar to ninein [M.musculus]
90. ESTs, Highly similar to PROTOPORPHYRINOGEN OXIDASE [H.sapiens]
91. ESTs, Highly similar to INSULIN-INDUCED GROWTH RESPONSE PROTEIN CL-6 [Rattus norvegicus]
92. ESTs, Highly similar to GLYCOGEN PHOSPHORYLASE, LIVER FORM [Homo sapiens]
93. ESTs
94. ESTs
95. ESTs
96. ESTs
97. ESTs
98. ESTs
99. ESTs
100. ESTs
101. ESTs
102. ESTs
103. ESTs
104. ESTs
105. ESTs
106. ESTs
107. ESTs
108. Pregnancy specific beta-1 glycoprotein 5
109. Human fibroblast growth factor homologous factor 1 (FHF-1) mRNA, complete cds
110. Homo sapiens mRNA for smallest subunit of ubiquinol-cytochrome c reductase, complete cds
111. HEAT SHOCK PROTEIN HSP 90-ALPHA
112. H.sapiens mRNA for protein-tyrosine-phosphatase (tissue type: foreskin)
113. GRO1 oncogene (melanoma growth stimulating activity, alpha)
114. INTERFERON-INDUCIBLE PROTEIN 1-8U
115. ESTs, Weakly similar to UTROPHIN [Homo sapiens]
116. ESTs, Weakly similar to kruppel-related zinc finger protein [H.sapiens]
117. ESTs, Moderately similar to Mouse 19.5 mRNA, complete cds [M.musculus]
118. ESTs, Highly similar to mitogen-induced [M.musculus]
119. Human extracellular protein (S1-5) mRNA, complete cds
120. PROBABLE TRANS-1,2-DIHYDROBENZENE-1,2-DIOL DEHYDROGENASE
121. ESTs
122. ESTs
123. PREGNANCY-SPECIFIC BETA-1 GLYCOPROTEIN D PRECURSOR

124. Human protease M mRNA, complete cds
125. Human mRNA for KIAA0146 gene, partial cds
126. Human cellular proto-oncogene (c-mer) mRNA, complete cds
127. Human TAR RNA binding protein (TRBP) mRNA, complete cds
128. Homo sapiens mRNA expressed in osteoblast, complete cds
129. Homo sapiens clone 22 mRNA, alternative splice variant alpha-1, complete cds
130. Epidermal growth factor receptor pathway substrate 15
131. JNK ACTIVATING KINASE 1
132. Human TFIIA gamma subunit mRNA, complete cds
133. ESTs, Moderately similar to nuclear LIM interactor [M.musculus]
134. ESTs, Highly similar to GLIA DERIVED NEXIN PRECURSOR [Homo sapiens]
135. ESTs
136. ESTs
137. ESTs
138. ESTs, Weakly similar to fractionated X-irradiation-induced 29 thymoma [M.musculus]
139. ESTs
140. ESTs
141. ESTs
142. ESTs
143. ESTs
144. ESTs
145. ESTs
146. ESTs
147. ESTs
148. ESTs
149. ESTs
150. ESTs
151. Integrin, alpha M (complement component receptor 3, alpha; also known as CD11b (p170), macrophage antigen alpha polypeptide)
152. Human pregnancy-specific beta-1 glycoprotein mRNA, complete cds
153. Human mRNA for KIAA0001 gene, complete cds
154. Human apM2 mRNA for GS2374 (unknown product specific to adipose tissue), complete cds
155. Human RalGDS-like 2 (RGL2) mRNA, partial cds
156. Homo sapiens hCPE-R mRNA for CPE-receptor, complete cds
157. Homo sapiens creatine transporter mRNA, complete cds
158. H.sapiens mRNA for hepatocyte nuclear factor 4 gamma
159. H.sapiens mRNA for Not56-like protein
160. H.sapiens mRNA for IL13 receptor
161. Cellular retinoic acid-binding protein [human, skin, mRNA, 735 nt]
162. Human hnRNP type A/B protein mRNA, complete cds
163. ESTs, Weakly similar to hTAFII100 [H.sapiens]
164. ESTs, Weakly similar to GAGE-4 protein [H.sapiens]
165. ESTs, Highly similar to CARCINOEMBRYONIC ANTIGEN CGM6 PRECURSOR [Homo sapiens]
166. ESTs, Weakly similar to ovary2 [D.melanogaster]

167. ESTs, Weakly similar to T-LYMPHOCYTE MATURATION-ASSOCIATED PROTEIN [H.sapiens]
168. ESTs, Highly similar to 8A-2V protein [M.musculus]
169. H.sapiens H4/g gene for H4 histone
170. Homo sapiens breast cancer-specific protein 1 (BCSG1) mRNA, complete cds
171. Human tetracycline transporter-like protein mRNA, complete cds
172. Homo sapiens (huc) mRNA, complete cds
173. ESTs
174. Solute carrier family 9 (sodium/hydrogen exchanger), isoform 1 (antiporter, Na+/H+, amiloride sensitive)
175. NKG2-D TYPE II INTEGRAL MEMBRANE PROTEIN
176. Matrix metalloproteinase 1 (interstitial collagenase)
177. MULTIDRUG RESISTANCE-ASSOCIATED PROTEIN 1
178. Human mRNA for KIAA0365 gene, partial cds
179. Human MHC class II HLA-DR2-Dw12 mRNA DQw1-beta, complete cds
180. CARTILAGE GLYCOPROTEIN-39 PRECURSOR
181. Aminolevulinate, delta-, synthase 2 (sideroblastic/hypochromic anemia)
182. Proteoglycan 1, secretory granule
183. Prolactin
184. Human growth hormone-dependent insulin-like growth factor-binding protein mRNA, complete cds
185. Human clone A9A2BRB6 (CAC)n/(GTG)n repeat-containing mRNA
186. Human GDP-dissociation inhibitor protein (Ly-GDI) mRNA, complete cds
187. ESTs
188. ESTs
189. ESTs, Weakly similar to putative p150 [H.sapiens]
190. ESTs, Weakly similar to HYPOTHETICAL 41.9 KD PROTEIN IN SDS3-THS1 INTERGENIC REGION [S.cerevisiae]
191. ESTs, Moderately similar to Lasp-1 protein [H.sapiens]
192. ESTs, Highly similar to MUSCLE-CADHERIN PRECURSOR [Mus musculus]
193. ESTs
194. ESTs
195. ESTs
196. ESTs
197. ESTs
198. ESTs
199. ESTs
200. ESTs
201. ESTs
202. ESTs
203. ESTs
204. ESTs
205. ESTs
206. Pregnancy-specific beta-1 glycoprotein 4
207. Phosphofructokinase, muscle
208. Human squamous cell carcinama of esophagus mRNA for GRB-7 SH2 domain protein, complete cds
209. Human mRNA for IgG Fc binding protein, complete cds

210. Human L-kynurenine hydrolase mRNA, complete cds
211. Human Ca2+-dependent activator protein for secretion mRNA, complete cds
212. Homo sapiens mRNA for low molecular mass ubiquinone-binding protein, complete cds
213. Homo sapiens mRNA for glutathione transferase A4-4
214. H.sapiens mRNA for laminin
215. Fatty acid binding protein 4, adipocyte
216. Cadherin 2, N-cadherin (neuronal)
217. CYSTATIN A
218. Prostacyclin-stimulating factor [human, cultured diploid fibroblast cells, mRNA, 1124 nt]
219. Integrin, beta 3 (platelet glycoprotein IIIa, antigen CD61)
220. ESTs, Weakly similar to putative p150 [H.sapiens]
221. ESTs, Highly similar to PROHIBITIN [Homo sapiens]
222. Homo sapiens mRNA from chromosome 5q21-22, clone:357Ex
223. ESTs, Weakly similar to putative p150 [H.sapiens]
224. ESTs, Weakly similar to T-complex protein 10A [H.sapiens]
225. ESTs, Weakly similar to Sxm1p [S.cerevisiae]
226. ESTs, Moderately similar to !!!! ALU SUBFAMILY SB2 WARNING ENTRY !!!! [H.sapiens]
227. ESTs, Moderately similar to PROTEIN-TYROSINE PHOSPHATASE [Autographa californica nuclear polyhedrosis virus]
228. Human (clone CTG-A4) mRNA sequence
229. Interleukin 1, alpha
230. Homo sapiens G protein beta 5 subunit mRNA, complete cds
231. Human homolog of yeast mutL (hPMS1) gene, complete cds
232. MELANOMA-ASSOCIATED ANTIGEN 4
233. Homo sapiens mRNA for hepatocyte growth factor activator inhibitor, complete cds
234. ESTs

TABLE 2

Association Between Expression Levels of Moesin, E-Cadherin, And Zyxin With Expression Levels of pRB or TP53.

Association With Rb Expression	Number Of Cases	*P-Value*

Moesin	158	0.005
E-Cadherin	153	0.003
Zyxin	151	0.001
Association With Tp53 Mutation	**Number Of Cases**	***P-Value***
Zyxin	65	0.03

EXAMPLE 2

Molecular Diagnosis And Outcome Prediction In Bladder Tumors By Gene Profiling Using cDNA Microarrays

This example discloses cDNA microarray analysis that demonstrates two main clusters segregating superficial from invasive transitional carcinomas. These clusters would provide prognostic information. Cytokeratin 20, neuropilin 2, p21 and p33ING1 were selected among the top ranked molecular targets differentially expressed between superficial and invasive tumors and validated by immunohistochemistry with tissue microarrays. Their expression patterns were associated with pathological stage, tumor grade, and altered RB expression. Moreover, p33ING1 expression levels were related to overall survival. Generation of a support vector machine algorithm revealed the relevance of WNT signaling and mitotic checkpoint alteration during bladder cancer progression. In summary, gene profiling successfully classified bladder tumors based on their histopathogenesis and clinical outcome, and identified molecular biomarkers of potential clinical significance.

Cell Lines And Tumor Samples For cDNA Analysis

Four bladder cancer cell lines: T24, J82, RT4, and HT1197 were obtained from ATCC (Rockville, MD) and maintained following standard procedures. All cells were grown and harvested at 75%-90% confluence no longer than 4-6 passages. Total RNA of cell lines was extracted using RNeasy (Qiagen, Valencia, CA). Fifteen patients with bladder cancer were included for the expression profiling

study. Specimens were collected under an IRB approved tissue procurement protocol. Bladder tumors embedded in OCT were macro-dissected to ensure a minimum of 75% of tumor cells. Total RNA from bladder tumors was isolated in two steps using TRIzol (Life Technologies, Carlsbad, CA), followed by RNeasy purification.

cDNA Microarrays Preparation And Image Acquisition

A set of 17,842 sequence-verified human IMAGE cDNA clones, representing both known genes and ESTs, were PCR amplified and spotted onto poly-lysine coated microscope slides by the Albert Einstein College of Medicine microarray facility (Cheung et al., 1999). Five micrograms of total RNA from each bladder tissue and pool of cell lines was linearly amplified using a single round (Hoos et al., 2001; Tseng et al., 2001). Amplified cRNA obtained from bladder tumors were labeled with Cy5 (red) and hybridized against amplified cRNA from the pool containing equal RNA quantities of the four cell lines labeled with Cy3 (green). Following hybridization, slides were washed, dried and scanned by an Axon automated laser scanner. GenePix software was used for gridding and signal intensities calculation (Tseng et al., 2001; Eisen et al., 1998).

Collection And Analysis of Data

Normalization: cDNA microarrays were normalized using an intensity-dependent algorithm (Tseng et al., 2001). Normalized fold changes in gene expression were then used to further analyze and cluster the bladder tumors (Eisen et al., 1998; Felsenstein, 1985).

Clustering: Before clustering, the data was filtered to select genes having both significant average intensities and fold changes removing the background of unchanging genes. It was required that genes have a fold change of at least 3 (up or down) and an average intensity greater than 316 for at least 2 samples using the geometric average of the two channel intensities. This filter reduced the number of genes from 17,842 to 15,609. The set of 15 bladder tumor samples were then analyzed using the hierarchical clustering with the Ward linkage method combined with non-parametric bootstrap resampling and consensus tree building to determine

the support for sample groupings (Eisen et al., 1998; Felsenstein, 1985).

Gene Ranking: Several scoring methods were applied to rank genes that could separate early-stage tumors from invasive organ-confined lesions and those developing metastatic bladder disease. Initially, the Mann-Whitney-Wilcoxon rank sum test was applied to identify genes differentially expressed between the two significant clusters (Dawson-Saunders and Trapp, 1994). Only genes showing a *p* value lower than 0.05 were considered for further analysis.

The method of single-variable logistic regression was applied labeling samples based on the histopathological records of the tumors analyzed in this study (Li and Yang, 2002; Xiong et al., 2001). If the log-expression level of gene j in sample i is x_{ji}, the sample is labeled as either one type of cancer (y_i=1) or another (y_i=0), the model of logistic regression is: Prob (y_i=1) = 1/(1+ exp(-a_j - b_j x_{ji})) where "exp" is the exponential function, and a_j and b_j are parameters in the model. The parameter values are obtained by the maximum likelihood estimation. Single variable logistic regression was carried out for all 15650 genes/ESTs according to the maximum likelihood. The decay of the maximum likelihood as a function of the rank is approximated a power-law function or Zipf's law (Li and Yang, 2002).

Multidimensional analysis was then performed taking the 5,616 genes providing data simultaneously in all the tumors. The gene-expression matrix (*N* experiments x *M* genes) was first pretreated. All columns (*M* genes) were re-normalized with the transformation $z = (x - \mu)/\sigma$, that is, a new matrix (X-matrix) was created where each column has a mean equal to zero and variance equal to one (*z*-transformation). The X-matrix was then studied with Q-mode Factor Analysis (FA) (Reyment and Joreskog, 1996; De las Rivas et al., 2002). FA seeks finding an underlying orthogonal factor model of the X-matrix of the form: $X = LF + E$, where L is the *loadings matrix,* F is the *scores matrix*, and E is the *residual matrix*. First, *loadings* were obtained by scaling the eigenvector matrix (P) obtained from Principal Components Analysis (PCA): $L = P\ L^{1/2}$. The optimal dimensionality was three, explaining 76.6 % of the variance. The factors were then rotated by means of a *varimax* rotation (Saitou & Nei, 1987; Reyment and Joreskog, 1996), and from these rotated factors the *scores* matrix was generated by an ordinary (unweighted) least squares procedure: $F = L^{-1/2}\ P^{*\prime}X$, where P* is the rotated eigenvectors, and the

prime denotes the transpose. These loading distances were clustered with a neighbor-joining (NJ) algorithm to build a dendrogram.

An estimation of the reliability of each branch was obtained by means of a jackknife bootstrap analysis. Bootstrap values were computed from selected random subsets of 75% of the genes and ESTs in the X-matrix, by reanalyzing the new resulting matrixes by FA and NJ. Relationships among samples obtained from the previous step were contrasted with the available clinical and pathological data. A supervised method (Cristiani & Shawe-Taylor, 2000) was applied in order to detect the combination of genes in the X-matrix that can optimally be able to explain these groups. Specifically, the hyperplane in the loadings space that optimally separates the previously defined groups of samples was obtained by means of a Support Vector Machine (SVM) algorithm. The FA scores were then projected onto the characteristic vector of this hyperplane and then were *z*-transformed. The sorted z-scores were used to select the important descriptors able to separate the groups. A z-score cutoff of 2 was used to obtain a subset of cDNAs with the best discriminant properties.

To obtain insights into the biochemical pathways involved in bladder tumor progression, the biological function index of the genes under study in GO was searched (Xie et al., 2002). Biological processes according to GO were available only for 1,044 genes out of the 5,616 under analysis. The inventors identified by t-test the genes differentially expressed among each of the four groups generated by FA-SVM, as well as among superficial lesions (pooling groups 1 and 2) and invasive tumors (pooling groups 3 and 4) versus the rest of the experiments. The average t-test is calculated within the different groups as a measure of enrichment of each biological process in the different groups. Only genes with t-test *p*-values lower than 0.01 were considered for further analysis. The statistical significance of the association of groups and biological processes (GO indexes) was evaluated by means of the hypergeometric distribution. The inventors computed the probability that at least x genes (with a t-test *p*-value lower than 0.01) were annotated within any given biological process in a random subset of n genes, where N denoted the total number of annotated genes (1044) in the entire dataset, and A the number of these genes with a particular annotation. These p-values were obtained according to the following equation:

$$p(x;N,A,n) = 1 - \sum_{i=0}^{x-1} \frac{\binom{A}{i}\binom{N-A}{n-i}}{\binom{N}{n}}, where: \binom{N}{n} = \frac{N!}{n!(N-n)!}$$

The following criteria were required for calculating the significance of the biological processes. Only biological processes (GO indexes) with more than 5 members in the set were selected for further analysis, and only annotated genes with t-test *p*-value lower than 0.01 within each biological process were considered. It was also required that at least two genes reached this significance within a biological process. Hypergeometric *p*-values lower than 0.05 were considered significant, but the inventors also focused on those marginally sub-optimals, to compensate the sparsity of GO annotations and the limitation of that non-annotated genes could be included in this analysis. Finally, significant cDNAs (according to a t-test), and belonging to significant or marginally significant GO pathways (*p*-value<0.01) were grouped and represented using a color-graded spectrum.

Clinical Validation of The Results

Tissue Samples In Tissue Microarrays Three different bladder cancer microarrays were used in this study (Hoos et al., 2001), including a total of 173 bladder primary transitional cell carcinoma (TCC) tumors obtained under IRB approved protocol. A total of 40 superficial and 64 invasive TCC tumors were analyzed in two microarrays. These tumors corresponds to grade 1 (n=24), grade 2 (n=8) and grade 3 (n=82) lesions. The third tissue microarray comprised a cohort of 69 bladder primary TCC cases with known p53/pRB status and annotated follow-up, including two superficial and 67 invasive lesions.

Immunohistochemistry Protein patterns of expression were assessed at the microanatomical level using both cytospins from cell lines studied (data not shown) and tissue microarrays outlined above. Standard avidin-biotin immunoperoxidase procedures were applied for immunohistochemistry. The following panel of mouse monoclonal antibodies were used: np-2 (clone 54; BD Transductions Labs, Lexington, KY); cytokeratin 20 (clone Ks20.8; DAKO, Denmark); cyclin E (clone cyE05; Neomarkers, Fremont, CA); p53 (clone 1801;

Calbiochem, Cambridge, MA); total pRB (clone 3C8; QED Bioscience, San Diego, CA); under-phosphorylated pRB (clone G99-549; BD Transductions Labs); ninjurin (clone 50; BD Transductions Labs); p33ING1 (clone CAB1; BD Transductions Labs); and p21/WAF1 (clone Ab-1, Calbiochem). Control tissues for specificity assessment were used according to manufacturers' recommendations. The inventors used a 20% cutoff for p53 staining, 10% for p21 and 25% for cyclin E. There is no consensus on the cutoffs of the immunohistochemical expression of the other markers, and thus they were analyzed as continuous variables, or taking several cutoffs when considered as categorical.

Statistical Analysis All TCC (n = 173) were used for the analysis of association between p53 and pRB with np-2, cytokeratin 20, cyclin E and p21. These cases were also utilized for evaluating marker expression versus histopathological stage and tumor grade, using the non-parametric Wilcoxon-Mann-Whitney and Kruskall-Wallis tests (Dawson-Saunders and Trapp, 1994). The consensus value of the representative cores from each tumor sample arrayed was used for statistical analyses.

The inventors analyzed the relationship of the cluster analysis of the bladder tumors to which expression profiling was performed with overall survival. Additionally, the association of the markers identified in the DNA microarray analysis to outcome was also evaluated using a subset of 69 cTCC cases for which follow up was available. Overall-survival time was defined as the months elapsed between transurethral resection or cystectomy and death from disease (or the last follow-up date). Patients who were alive at the last follow-up or lost to follow-up were censored. For survival analysis, bootstrapping cluster and biomarkers were analyzed as categorical variables. The association of the marker expression levels with overall survival was analyzed using the Wald test, and the log-rank test was used to examine their relationship when different cutoffs were applied. Survival curves were plotted using Kaplan-Meier methodology. Associations between markers were analyzed using Kendall's tau b test using the SPSS statistical package (version 8.0).

Molecular Classification And Predictive Value of Hierarchical Clustering

The present analysis was carried out on the basis of two

complementary sets of experiments. Initially, bladder tumors were analyzed using cDNA microarrays to identify differentially expressed genes among histopathologically distinct tumors (Table 3). The transcriptome of 15 bladder tumors was compared against a pool of four bladder cancer cell lines containing equal RNA quantities using cDNA microarrays with 17,842 known genes and expressed sequence tags (ESTs) (Cheung et al., 1999). Secondly, the potential clinical significance of the selected targets identified by cDNA microarrays was validated at the microanatomical level using immunohistochemistry on tissue microarrays containing well-characterized bladder carcinomas (Hoos et al., 2001). A cohort of superficial and invasive bladder neoplasms was used to evaluate the association between biomarkers and histopathological stage and grade (n=173). A subset of these bladder tumors (n=69), with characterized p53 and pRB alterations and clinical follow-up, was used to delineate associations between potential novel biomarkers and these cell cycle regulators as well as with patient outcome.

The use of unsupervised hierarchical clustering combined with non-parametric bootstrap analysis classified primary bladder carcinomas based on their histopathological criteria. Overall, the superficial tumors clustered together and were segregated from invasive transitional carcinomas. Cases developing metastasis and displaying a shorter survival could be distinguished from others displaying a longer survival and those that had organ-confined disease. The bootstrap resampling technique was able to establish a high confidence of these clusters (Figure 6A). Patients whose tumor samples were subjected to gene profiling had a median follow-up of 12 months (mean: 14.3 months, range: 1 to 44 months) (Table 3). The superficial and the invasive clusters were significantly associated with overall survival ($p=0.0025$) (Figure 6B). These results revealed the diagnostic and prognostic utility of unsupervised clustering since the identified clusters were associated with histopathogenesis and overall survival.

A multidimensional analysis of 4729 genes providing expression data in all the tumors, revealed four groups of expression profiles consistent with the previously identified clusters (Figure 6C, Table 3). Interestingly, this additional analysis also revealed that the gene expression profiles of certain superficial tumors were more similar to some organ-confined invasive lesions (Figure 6D).

Identification of Genes Differentially Expressed Between Superficial And Invasive Bladder Cancer

Gene identification was the next step in the study. Several scoring methods were applied to rank the genes according to their ability to separate the superficial from the invasive organ-confined and those developing metastatic bladder tumors (Tables 4-6). First, the Mann-Whitney-Wilcoxon rank sum test was applied as a standard means for gene identification between two groups. The goal was to identify genes differentially expressed between the superficial and invasive clusters. It was observed that the first 120 genes correctly classified the samples contained in each cluster ($p=0.033$) (Table 4). Two genes, p21 and cyclin E were selected for further study due to their participation in the p53 and RB signaling pathways, both of which are frequently altered in bladder cancer progression.

A single-gene variable logistic regression analysis was carried out as a standard classification/discrimination model to rank genes by their classification performance. The goal was to identify genes differentially expressed between superficial and the invasive non-metastatic together with invasive tumors associated with development of metastatic disease. The results demonstrated that any of the 92 top-ranked genes could differentiate superficial versus invasive tumors based on their levels of gene expression. For those genes ranked from 93 to 500 using this analysis, a maximum of 3 misclassifications could be obtained. The inventors chose to focus on the 92 genes that provided no misclassification among superficial versus invasive and those developing metastasis (Table 5). Two genes from this initial group, cytokeratin 20 and neuropilin-2 (np-2), coding for soluble proteins with potential role for tumor marker development were selected and studied by immunohistochemistry on tissue microarrays.

Finally, the inventors continued to further elucidate the genes that best characterized each of the four groups generated by the multidimensional approach (Table 6). Overall, groups 2 and 4 in this analysis included tumors with the worst clinical outcome. Since the inventors were interested in the identification of genes with potential prognostic utility, one gene from each of these groups (ninjurin and p33ING1) was selected in order to evaluate their potential prognostic value.

Association of The Identified Markers With Tumor Stage, Grade, And P53/ RB Expression And Overall Survival

The potential roles of the identified target genes in diagnosing patients with superficial and invasive disease were analyzed using tissue microarrays containing transitional cell carcinoma (TCC) of different stages and grades. Overall, the results demonstrated that the expression of these molecular markers was associated with tumor stage, grade, p53/pRB expression and overall survival. Cytokeratin 20, np-2, p21, and p33ING1 were differentially expressed in superficial and invasive tumors. In the subset of patients analyzed, there was a significant correlation between the expression of these proteins and tumor stage and grade (Table 7a). Levels of p33ING1 expression were easily detectable in normal urothelium, and in the majority of superficial TCC, but were much lower in invasive tumors (Figure 7).

Since both p53 and RB signaling pathways are frequently altered during bladder cancer progression, the association of these biomarkers with p53 and pRB status was also evaluated. The expression of cyclin E was significantly associated with p53 expression; cytokeratin 20, np-2, p21, and p33ING1 were all associated with altered pRB expression (Table 7b). A significant correlation between p33ING1 expression with cyclin E and p21 was also noted.

When the overall survival prognostic utility of cytokeratin 20, np-2, p21 and cyclin E was evaluated using 69 TCC for which clinical follow-up was available, it was observed that only the expression of p33ING1 was significantly associated with overall survival ($p=0.02$). Patients displaying a higher expression of p33ING1 showed a shorter survival than those with low expression of this protein (Figure 7C).

Molecular Pathways Involved In Bladder Cancer

A supervised method, based on a Support Vector Machine algorithm (SVM), was applied to detect the combination of genes that can optimally explain the two main groups (superficial from invasive transitional carcinomas) identified by multidimensional analysis. The goal was to obtain insight about the biochemical pathways involved in tumor progression. The subset of genes with the most optimal

discriminatory properties was clustered using the correlation coefficient of the original data as the similarity measure. These genes were then grouped according to the molecular pathways in which they are involved. SVM revealed that WNT signaling pathway and mitotic spindle checkpoint are among the most important networks altered during bladder cancer progression (Figure 8). Representative genes of these pathways that were top-ranked based on their z-score included FAK, zyxin or cdc16.

TABLE 3

Clinical, Histopathological And Epidemiological Characteristics of Bladder Cancer Patients

Patient ID	BT	MD	Age	Sex	TNM	Carcinoma In situ	Prostate Cancer	Smoking Habit	Familiar Cancer History	Follow-up (months)	Clinical Outcome
174	1	1	67	M	TISG3N0	YES	YES	NO	NO	12	NED
160	1	1	83	M	TAG1N0	YES	YES	NO	NO	20	NED
157	1	1	61	M	TISG3N0	YES	NO	NO	NO	44	NED
134	1	1	80	M	T3BG3NO T4G3N0M	YES	YES	YES	YES	13	NED
170	2	2	55	M	1	YES	YES	YES	NK	4	DOD
168	2	2	61	M	T4BN1M2	YES	YES	YES	NK	1	DOD
169	1	3	75	F	TAG3NO	NO	NO	NO	NK	41	NED
165	1	3	75	M	TAG1NO	NO	NO	NO	NO	11	NED
163	1	3	65	M	TAG1NO	NO	NO	YES	YES	17	NED
162	1	3	60	M	TISG3NO	YES	YES	YES	YES	15	NED
141	2	4	49	F	TAG3M10	NO	NO	YES	YES	3	DOD
135	2	4	64	F	T4BG3N1	YES	NO	YES	YES	11	DOD
133	2	4	83	M	T3BG3NO	YES	YES	YES	YES	1	DOD
130	2	4	72	M	T3BG3N0	NO	YES	NO	NO	13	NED
124	2	4	59	F	T3AN1M0	NO	NO	YES	NO	9	NED

BT: Bootstrap Clustering; MD: Multidimensional Grouping; TNM: Tumor Node Metastases; Clinical outcome: NED (no evidence of disease), DOD (death of disease).

TABLE 4

120 Genes Identified By Mann-Whitney-Wilcoxon Test

Acc. Num	Name
AA011414	Homo sapiens fibrinogen alpha chain preproprotein (FGA) gene, complete cds, alternatively spliced.789 (0.0)
AA021434	Homo sapiens, Similar to retinal degeneration B beta, clone MGC:14375 IMAGE:4299595, mRNA, complete cds.325 (8e-87). Human DNA sequence from clone RP5-973N23 on chromosome 6p12.3-21.2, complete sequence 42 (0.18)
AA021464	Homo sapiens chromosome 8 clone RP11-86O15, complete sequence. 278 (4e-72)
AA028884	Homo sapiens clone RP4-647J21, complete sequence.914 (0.0)
AA034115	Homo sapiens 12q BAC RP11-415I12 (Roswell Park Cancer Institute Human BAC Library) complete sequence.967 (0.0)
AA035095	Homo sapiens, Similar to Breakpoint cluster region protein, uterine leiomyoma, 1; barrier to autointegration factor, clone IMAGE:3027737, mRNA. Homo sapiens, Similar to Breakpoint cluster region protein, uterine leiomyoma, 1; barrier to autointegration factor, clone MGC:14564 IMAGE:4074168, mRNA, complete cds. Homo sapiens Breakpoint cluster region protein, uterine leiomyoma,1; barrier to autointegration factor BCRP1), mRNA.779 (0.0)
AA043806	Human beta 3-endonexin mRNA, long form and short form, complete cds.884 (0.0)Homo sapiens integrin beta 3 binding protein (beta3-endonexin) (ITGB3BP), mRNA.761 (0.0)Homo sapiens nuclear receptor co-activator NRIF3 (NRIF3) mRNA, alternatively spliced, complete cds.726 (0.0)
AA074666	Human DNA sequence from clone CTB-1189H8 on chromosome 1, complete sequence.121 (e-25)
AA083385	Homo sapiens BTB/POZ domain containing 1 protein (BTBD1) mRNA, complete cds.575 (e-161)
AA101348	Homo sapiens, clone IMAGE:4025624, mRNA.Homo sapiens similar to dendritic cell protein (LOC63319), mRNA. 809 (0.0)Homo sapiens dendritic cell protein (GA17), mRNA.801 (0.0)
AA127058	Homo sapiens similar to RIKEN cDNA 2610103J23 gene (LOC92140), mRNA. Homo sapiens genomic DNA, chromosome 8q23, clone: KB1907C4.535 (E-150)
AA132065	Homo sapiens chromosome 5 clone CTC-222O22, complete sequence. Homo sapiens mRNA for SMAP-5, partial cds.383 (e-103)Human DNA sequence from clone RP1-315G1 on chromosome Xq24-25. Contains a PDZ (DHR, GLGF) domain protein pseudogene, the API3 gene for apoptosis inhibitor 3 (XIAP, HILP), a putative novel gene, ESTs, STSs, GSSs and a putative CpG island, complete sequence.44 (0.15)
AA143509	Homo sapiens pyrroline-5-carboxylate synthetase (glutamate gamma-semialdehyde synthetase) (PYCS), mRNA.993 (0.0)Human DNA sequence from clone RP11-7D5 on chromosome 10, complete sequence.696 (0.0)
AA147928	Human DNA sequence from clone RP4-758J18 on chromosome 1p36.31-36.33, complete sequence.999 (0.0)
AA156863	Homo sapiens phosphomannomutase mRNA, complete cds.1124 (0.0)
AA165403	Homo sapiens BM-019 mRNA, complete cds. Homo sapiens acid cluster protein 33 (ACP33), mRNA.Homo sapiens GL010 mRNA, complete cds.975 (0.0)
AA172210	Homo sapiens cDNA FLJ30082 fis, clone BGGI12000839. 327 (2e-87)
AA190401	Homo sapiens, Similar to esterase 10, clone IMAGE:3350277, mRNA.396 (e-108)Homo sapiens chromosome 22q11 clone p1087I10, complete sequence.Homo sapiens, BH3 interacting domain death agonist, clone MGC:15319 IMAGE:4025880, mRNA, complete cds.394 (e-107)
AA256462	Homo sapiens, nudix (nucleoside diphosphate linked moiety X)-type motif 3, clone MGC:12752 IMAGE:4303483, mRNA, complete cds.404 (e-110)
AA279188	Homo sapiens a disintegrin and metalloproteinase domain 8 (ADAM8), mRNA.823 (0.0)Homo sapiens mRNA for transmembrane protein, complete cds.783 (0.0)

AA394148	
AA402766	Homo sapiens small membrane protein 1 (SMP1) gene, complete cds.565 (e-158)H.sapiens mRNA for rhesus polypeptide (RhVI).103 (e-19)
AA421518	Homo sapiens adaptor-related protein complex 2, sigma 1 subunit (AP2S1), mRNA. 862 (0.0). Homo sapiens adaptor-related protein complex 2, sigma 1 subunit (AP2S1), transcript variant AP17, mRNA.H.sapiens mRNS for clathrin-associated protein.805 (0.0) Homo sapiens adaptor-related protein complex 2, sigma 1 subunit (AP2S1), transcript variant AP17delta, mRNA.799 (0.0)
AA424578	Homo sapiens cDNA FLJ32289 fis, clone PROST2000432, highly similar to Homo sapiens mRNA for UDP-Gal:GlcNAc galactosyltransferase.833 (0.0) Homo sapiens, UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 3, clone MGC:12774 IMAGE:3677118, mRNA, complete cds.825 (0.0)
AA425861	Homo sapiens, clone MGC:27221 IMAGE:4306900, mRNA, complete cds.Homo sapiens chromosome 19 clone LLNLF-172E10, complete sequence.507 (e-141) Homo sapiens similar to peroxisomal enoyl-coenzyme A hydratase-like protein; delta3,5-delta2,4-dienoyl-CoA isomerase; peroxisomal enoyl-CoA hydratase 1; dienoyl-CoA isomerase (LOC115289), mRNA.502 (e-140)
AA430520	Homo sapiens, CDP-diacylglycerol--inositol 3-phosphatidyltransferase (phosphatidylinositol synthase), clone MGC:1328 IMAGE:3139134, mRNA, complete cds.844 (0.0)Homo sapiens CDP-diacylglycerol--inositol 3-phosphatidyltransferase (phosphatidylinositol synthase) (CDIPT), mRNA.837 (0.0)Homo sapiens phosphatidylinositol synthase (PIS) mRNA, complete cds. 827 (0.0)
AA434068	Homo sapiens TRF2-interacting telomeric RAP1 protein (RAP1), mRNA.809 (0.0)
AA446453	Homo sapiens similar to PREFOLDIN SUBUNIT 5 (C-MYC BINDING PROTEIN MM-1) (MYC MODULATOR 1) (H. sapiens) (LOC121342), mRNA.Homo sapiens, prefoldin 5, clone MGC:5329 IMAGE:2900793, mRNA, complete cds.Homo sapiens mRNA for MM-1 alpha, complete cds.Homo sapiens mRNA for c-myc binding protein, complete cds.797 (0.0)
AA447696	Homo sapiens CGI-127 protein (LOC51646), mRNA.739 (0.0)
AA449831	Homo sapiens, growth factor receptor-bound protein 2, clone MGC:1737 IMAGE:3345524, mRNA, complete cds.Homo sapiens growth factor receptor-bound protein 2 (GRB2), mRNA.Homo sapiens epidermal growth factor receptor-binding protein GRB2 (EGFRBP-GRB2) mRNA sequence.492 (e-136)
AA450227	Homo sapiens proteasome (prosome, macropain) 26S subunit,non-ATPase, 4 (PSMD4), mRNA.Homo sapiens, proteasome (prosome, macropain) 26S subunit, non-ATPase, 4, clone MGC:8410 IMAGE:2820813, mRNA, complete cds.Human antisecretory factor-1 mRNA, complete cds. 345 (9e-93)
AA450265	Homo sapiens proliferating cell nuclear antigen (PCNA), mRNA. Human DNA sequence from clone RP4-746J20 on chromosome 20. Contains the PCNA gene for proliferating cell nuclear antigen, the 5' end of the gene HSPC274, the 5' end of the CDS2 gene for CDP-diacylglycerol synthase (phosphatidate cytidylyltransferase) 2,ESTs, STSs, GSSs and three CpG islands, complete sequence.728 (0.0) Human cyclin protein gene, complete cds. 690 (0.0)
AA454566	Homo sapiens similar to putative (H. sapiens) (LOC120036), mRNA.418 (e-114)
AA454862	Homo sapiens CGI-135 protein (LOC51024), mRNA.509 (e-142)
AA455150	Human chromosome 14 DNA sequence BAC C-2011M8 of library CalTech-D from chromosome 14 of Homo sapiens (Human), complete sequence.387 (e-105)
AA455281	Homo sapiens, defender against cell death 1, clone MGC:17117 IMAGE:3454611, mRNA, complete cds.934 (0.0)Human mRNA for DAD-1, complete cds.898 (0.0)
AA456136	Homo sapiens genomic DNA, chromosome 8q23, clone: KB1460A1.852 (0.0)
AA457092	
AA457162	Homo sapiens similar to RIKEN cDNA 2810417J12 gene (LOC114984), mRNA.Homo sapiens chromosome 16, cosmid clone 399H11 (LANL), complete sequence. 353 (6E-95)

AA457725	Homo sapiens chromosome 17, clone hRPC.4_G_17, complete sequence.Homo sapiens GABA(A) receptor-associated protein (GABARAP), mRNA.Homo sapiens ganglioside expression factor 2 homolog mRNA, complete cds.783 (0.0)Homo sapiens MM46 mRNA, complete cds. Homo sapiens FLC3B mRNA for MAP1 light chain 3 related protein, complete cds.767 (0.0)
AA458661	
AA459663	Homo sapiens peroxiredoxin 4 (PRDX4), mRNA.Homo sapiens, thioredoxin peroxidase (antioxidant enzyme), clone MGC:22910 IMAGE:4075326, mRNA, complete cds. Human antioxidant enzyme AOE37-2 mRNA, complete cds.585 (e-165)
AA464152	Human DNA sequence from clone RP11-502H18 on chromosome 1, complete sequence.702 (0.0) Homo sapiens quiescin Q6 (QSCN6), mRNA.700 (0.0) Homo sapiens bone-derived growth factor (BPGF-1) mRNA, complete cds.652 (0.0)Human chorionic gonadotropin beta subunit mRNA, 5' flank, clone pCG-beta-474.92 (4e-16)
AA464192	Homo sapiens, Similar to hypothetical protein, clone MGC:3404 IMAGE:3530647, mRNA, complete cds.Homo sapiens HSPC227 mRNA, complete cds.Human DNA sequence from clone RP13-26D14 on chromosome Xq13.2-21.1, complete sequence.848 (0.0)
AA465031	Homo sapiens, pleckstrin homology, Sec7 and coiled/coil domains 2 (cytohesin-2), clone MGC:642 IMAGE:3538580, mRNA, complete cds.Homo sapiens cytohesin-2 mRNA, complete cds.Human Sec7p-like protein mRNA, partial cds.1003 (0.0)Homo sapiens pleckstrin homology, Sec7 and coiled/coil domains 2 (cytohesin-2) (PSCD2), mRNA.987 (0.0)H.sapiens mRNA for Arno protein.967 (0.0)
AA465378	Homo sapiens cDNA FLJ33055 fis, clone TRACH1000125, highly similar to Ig delta chain.601 (e-169)Human immunoglobulin C(mu) and C(delta) heavy chain genes (constant regions).Human Ig germline delta H-chain C-region gene, C-delta-3 domain (CLL lymphocyte).418 (e-114)
AA465593	Homo sapiens proteasome (prosome, macropain) subunit, alpha type, 3 (PSMA3), mRNA. Human mRNA for proteasome subunit HC8. 787 (0.0)
AA478268	Homo sapiens, Similar to C-terminal binding protein 1, clone MGC:12707 IMAGE:4128336, mRNA, complete cds.414 (E-113)Homo sapiens C-terminal binding protein 1 (CTBP1), mRNA.Homo sapiens phosphoprotein CtBP mRNA, complete cds.387 (E-105)
AA485052	Homo sapiens, proteasome (prosome, macropain) 26S subunit, non-ATPase, 3, clone MGC:9893 IMAGE:3868681, mRNA, complete cds. Homo sapiens clone 308 proteasome subunit p58 mRNA, complete cds.236 (4e-60)
AA486313	Homo sapiens low density lipoprotein-related protein-associated protein 1 (alpha-2-macroglobulin receptor-associated protein 1) (LRPAP1), mRNA. Human DNA sequence from cosmid L98A6, Huntington's Disease Region, chromosome 4p16.3 contains LRPAP1(low-density lipoprotein-associated protein-1) and CpG islands. Human alpha-2-macroglobulin receptor-associated protein mRNA, complete cds.593 (e-167)
AA486374	Homo sapiens lysyl-tRNA synthetase mRNA, complete cds; nuclear gene for mitochondrial product; alternatively spliced.Homo sapiens lysyl-tRNA synthetase (KARS), mRNA.Human mRNA for KIAA0070 gene, partial cds.769 (0.0)
AA486761	Homo sapiens tyrosyl-tRNA synthetase (YARS), mRNA.440 (E-121)Homo sapiens EGF-like-domain, multiple 4 (EGFL4), mRNA.40 (0.83)
AA487020	Homo sapiens isoprenylcysteine carboxyl methyltransferase (ICMT), mRNA. Human DNA sequence from clone RP1-120G22 on chromosome 1p36.21-36.33, complete sequence.398 (e-109)Homo sapiens MSTP098 (MST098) mRNA, complete cds.Homo sapiens prenylcysteine carboxyl methyltransferase (PCCMT) mRNA, complete cds.391 (e-106)
AA487223	Homo sapiens synovial sarcoma translocation gene on chromosome 18-like 2 (SS18L2), mRNAHomo sapiens kiaa-iso protein mRNA, complete cds..515 (E-144)Homo sapiens, synovial sarcoma translocation gene on chromosome 18-like 2, clone MGC:22369 IMAGE:4658643, mRNA, complete cds.511 (E-143)

AA487265	Homo sapiens KIAA0102 gene product (KIAA0102), mRNA.730 (0.0) Homo sapiens similar to MICROSOMAL SIGNAL PEPTIDASE 25 KDA SUBUNIT (SPC25) (LOC94390), mRNA. Human DNA sequence from clone RP5-1053E7 on chromosome 1p21.1-21.3, complete sequence.722 (0.0)
AA487899	Homo sapiens, uncharacterized hematopoietic stem/progenitor cells protein MDS027, clone MGC:17655 IMAGE:3858231, mRNA, complete cds.Homo sapiens uncharacterized hematopoietic stem/progenitor cells protein MDS027 mRNA, complete cds.232 (5e-59)
AA489400	Homo sapiens proteasome (prosome, macropain) subunit, beta type, 7(PSMB7), mRNA.Human mRNA for proteasome subunit z, complete cds.878 (0.0)Homo sapiens, proteasome (prosome, macropain) subunit, beta type, 7, clone IMAGE:3912276, mRNA, partial cds.870 (0.0)
AA490047	Homo sapiens poly(rC) binding protein 1 (PCBP1), mRNA.Human alpha-CP1 mRNA, complete cds.H.sapiens hnRNP-E1 mRNA.1011 (0.0) H.sapiens mRNA for nucleic acid binding protein sub2.3.987 (0.0)
AA490390	Homo sapiens small acidic protein (IMAGE145052), mRNA.Homo sapiens chromosome 11, clone RP11-4B7, complete sequence.216 (8e-54)
AA496359	Homo sapiens similar to immediate early protein (LOC95255), mRNA.Homo sapiens immediate early protein (ETR101), mRNA.Homo sapiens, Similar to kinesin family member 5B, clone MGC:15265 IMAGE:4297793, mRNA, complete cds.208 (6e-52)
AA496784	Homo sapiens, clone IMAGE:3959959, mRNA, partial cds. Human (chromosome 3p25) membrane protein mRNA.1015 (0.0)Homo sapiens SEC13-like 1 (S. cerevisiae) (SEC13L1), mRNA.1007 (0.0)
AA496948	Homo sapiens M5-14 protein (LOC51300), mRNA.636 (e-180)
AA504128	Human DNA sequence from clone RP4-800J21 on chromosome 20 Contains ESTs, STSs, GSSs and two CpG islands. Contains the 3' part of the RAE1 gene for a homolog to RNA export protein 1 from S.pombe and the gene for the ssDNA binding protein SEB4D (HSRNASEB).n, complete sequence.648 (0.0)Homo sapiens RAE1 (RNA export 1, S.pombe) homolog (RAE1), mRNA. 630 (e-178)
AA504617	Homo sapiens RNA-binding protein (autoantigenic) (RALY), transcript variant 1, mRNA.Homo sapiens heterogeneous nuclear ribonucleoprotein, alternate transcript (RALY) mRNA, complete cds.256 (7e-66) Homo sapiens autoantigen p542 mRNA, complete cds. 244 (3e-62)
AA598759	Homo sapiens, phosphogluconate dehydrogenase, clone MGC:8331 IMAGE:2819330, mRNA, complete cds.571 (e-160)
AA598815	Homo sapiens proteasome (prosome, macropain) subunit, alpha type, 5 (PSMA5), mRNA.H.sapiens mRNA for macropain subunit zeta.902 (0.0)
AA620479	Human DNA sequence from clone RP11-165N19 on chromosome 9, complete sequence.313 (3e-83)
AA625981	Homo sapiens FK506 binding protein 1A (12kD) (FKBP1A), transcript variant 12B, mRNA.Homo sapiens, tubulin, beta 5, clone MGC:4029 IMAGE:3617988, mRNA, complete cds.805 (0.0)
AA629584	Homo sapiens ADP-ribosylation factor 5 (ARF5), mRNA.Human ADP-ribosylation factor (hARF5) mRNA, complete cds. 668 (0.0)
AA633757	Homo sapiens splicing factor 3b, subunit 2, 145kD (SF3B2), mRNA. Human spliceosome associated protein (SAP 145) mRNA, complete cds.682 (0.0)
AA669341	Homo sapiens, unactive progesterone receptor, 23 kD, clone MGC:4004 IMAGE:2821965, mRNA, complete cds.775 (0.0)Mus musculus telomerase binding protein, p23 (Tebp-pending), mRNA.331 (4e-48)
AA680322	Homo sapiens clone RP5-855F16, complete sequence. Homo sapiens NADH:ubiquinone oxidoreductase MLRQ subunit (NDUFA4) mRNA, complete cds.353 (3e-95)
AA682613	Homo sapiens, craniofacial development protein 1, (CFDP1), clone MGC:5126 IMAGE:3449836, mRNA, complete cds. Homo sapiens BCNT mRNA, complete cds. Homo sapiens mRNA for p97 homologous protein, partial cds.819 (0.0)
AA683085	

AA705886	Homo sapiens MAX interacting protein 1 (MXI1), mRNA. Homo sapiens, Similar to MAX-interacting protein 1, clone MGC:9999 IMAGE:3882557, mRNA, complete cds.700 (0.0)
AA775415	Homo sapiens SMT3 suppressor of mif two 3 homolog 2 (yeast) (SMT3H2), mRNA.Homo sapiens similar to SMT3 (suppressor of mif two 3, yeast) homolog 2 (LOC91930), mRNA.H.sapiens mRNA for SMT3B protein. 498 (e-138)
AA862434	
AA934762	Homo sapiens, proteasome (prosome, macropain) 26S subunit, non-ATPase, 11, clone MGC:8396 IMAGE:2820583, mRNA, complete cds.Homo sapiens mRNA for 26S proteasome subunit p44.5, complete cds.710 (0.0)
AA935560	Homo sapiens relaxin 2 (H2) (RLN2), mRNA. Human DNA sequence from clone RP11-12D24 on chromosome 9p23-24.3 Contains the RLN1 gene encoding two isoforms of Relaxin 1(H1),the RLN2 gene encoding two isoforms of Relaxin 2 (H2), a putative novel gene, a HGM17 (high-mobility group (non-histone chromosomal) protein 17) pseudogene, four CpG island, ESTs STSs and GSSs, complete sequence.Human mRNA for prepro-relaxin H2.367 (5e-99)
AI017703	Homo sapiens, eukaryotic translation initiation factor 3, subunit 3 (gamma, 40kD), clone MGC:8431 IMAGE:2821133, mRNA, complete cds.914(0.0)Homo sapiens eukaryotic translation initiation factor 3, subunit 3 (gamma, 40kD) (EIF3S3), mRNA.Homo sapiens translation initiation factor eIF3 p40 subunit mRNA,complete cds.894 (0.0)
H05769	Homo sapiens, clone MGC:5564, mRNA, complete cds. Human DNA sequence from clone 108K11 on chromosome 6p21 Contains SRP20 (SR protein family member), Ndr protein kinase gene similar to yeast suppressor protein SRP40, EST and GSS, complete sequence.525 (e-147)
H17158	Homo sapiens ring finger protein 4 (RNF4), mRNA.448 (e-123)
H20652	Human mRNA for KIAA0069 gene, partial cds.712 (0.0)Homo sapiens ADP-ribosylation factor-like 6 interacting protein (ARL6IP), mRNA. 698 (0.0)
H21040	Human chromosome 14 DNA sequence BAC R-241N4 of library RPCI-11 from chromosome 14 of Homo sapiens (Human), complete sequence.381 (e-103)
H23366	
H23880	Homo sapiens hypothetical protein MGC14327 (MGC14327), mRNA.551 (e-154)
H54093	Homo sapiens mRNA for KIAA1470 protein, partial cds.496 (e-138)
H73731	Homo sapiens mRNA for KIAA0601 protein, partial cds.347 (5e-93)Human DNA sequence from clone RP1-184J9 on chromosome 1p35.1-36.12, complete sequence.339 (e-90)
H84444	Homo sapiens similar to CG4332 gene product (H. sapiens) (LOC134180), mRNA.Homo sapiens cDNA FLJ14400 fis, clone HEMBA1003742, weakly similar to Homo sapiens cleft lip and palate transmembrane protein 1 (CLPTM1) mRNA.Homo sapiens CRR9 mRNA for cisplatin resistance related protein CRR9p, complete cds.668 (0.0)
H93463	Human DNA sequence from clone RP11-108F13 on chromosome 1, complete sequence.143 9e-31)
H94897	Homo sapiens glycosyltransferase AD-017 (AD-017), mRNA.696 (0.0)Human chromosome 3p21.1 gene sequence.454 (E-125)
H99502	Homo sapiens hypothetical protein MGC2941 (MGC2941), mRNA.607 (e-171)Homo sapiens, Similar to RIKEN cDNA 2410141M05 gene, clone IMAGE:4155987, mRNA, partial cds.Homo sapiens chromosome 17, clone hRPC.4_G_17, complete sequence.605 (e-170)
N30811	ESTS
N54338	Homo sapiens, clone MGC:20354 IMAGE:4548461, mRNA, complete cds.Homo sapiens mRNA full length insert cDNA clone EUROIMAGE 2176457.801 (0.0)

N69283	Homo sapiens, TAR DNA binding protein, clone MGC:1433 IMAGE:3506121, mRNA, complete cds.Homo sapiens TAR DNA binding protein (TARDBP), mRNA. Homo sapiens MASP-2 gene (partial), PM-scl gene, FRAP2 gene and CDT6 gene, clone RPCI.11-99P18.Human DNA sequence from clone RP4-635E18 on chromosome 1p36.11-36.31, complete sequence.688 (0.0)
N73536	Homo sapiens hypothetical protein (BM-002), mRNA.Homo sapiens PRO3033 mRNA, complete cds.496 (e-138)
N91962	Homo sapiens, eukaryotic translation elongation factor 1 epsilon 1, (EEF1E1),clone MGC:12352 IMAGE:3685030, mRNA, complete cds. Homo sapiens p18 protein mRNA, complete cds.Homo sapiens mRNA for p18 component of aminoacyl-tRNA synthetase complex, complete cds.309 (4e-82)
R16165	Homo sapiens hypothetical PHD zinc finger protein XAP135 gene, complete cds.Human DNA sequence from clone RP11-160E12 on chromosome 6, complete sequence.716 (0.0)
R22439	Homo sapiens, transmembrane protein 4, clone MGC:1545 IMAGE:3344788, mRNA, complete cds. 640 (0.0)Homo sapiens putative secreted protein ZSIG9 (ZSIG9) mRNA, complete cds.622 (e-176) Homo sapiens transmembrane protein 4 (TMEM4), mRNA.80 (2e-12)
R24543	Homo sapiens neuroepithelial cell transforming gene 1 (NET1), mRNA.Human guanine nucleotide regulatory protein (NET1) mRNA, complete cds. Homo sapiens mRNA for Rho guanine nucleotide-exchange factor, splice variant NET1A.283 (6e-74)
R25377	Homo sapiens DEK oncogene (DNA binding) (DEK), mRNA.Homo sapiens partial unknown mRNA from drug-resistant melanoma cells, 3'UTR, clone DSM-4.Human DNA sequence from clone 298J15 on chromosome 6p22.3-23 Contains dek (putative oncogene), EST, GSS, CA repeat, CpG island, complete sequence.H.sapiens dek mRNA.492 (E-136)
R27552	Human chromosome 14 DNA sequence BAC R-354E14 of library RPCI-11 from chromosome 14 of Homo sapiens (Human), complete sequence.Homo sapiens mRNA for KIAA1333 protein, partial cds.428 (E-117)
R49144	Homo sapiens cDNA FLJ30169 fis, clone BRACE2000864, highly similar to TUBULIN ALPHA-4 CHAIN.349 (7e-94)Human HALPHA44 gene for alpha-tubulin, exons 1-3.339 (e-89)
R53889	Homo sapiens high-mobility group (nonhistone chromosomal) protein 14 (HMG14), mRNA. Human non-histone chromosomal protein HMG-14 mRNA, complete cds.204 (5e-50)
R55763	Human DNA sequence from clone RP1-198I9 on chromosome 6q12-13. Contains the gene KIAA1411, ESTs, STSs and GSSs, complete sequence.486 (e-135)
R69307	Homo sapiens leucine aminopeptidase (LOC51056), mRNA.474 (e-131)Human p21 (WAF1) gene, partial promoter sequence.381 (e-103)
R76314	Homo sapiens ras homolog gene family, member G (rho G) (ARHG), mRNA. 672 (0.0)
R78514	Homo sapiens, Similar to CG11985 gene product, clone MGC:3133 IMAGE:3352960, mRNA, complete cds. Homo sapiens hypothetical protein MGC3133 (MGC3133), mRNA. Human DNA sequence from clone 197L1 on chromosome 6q24.1-25.2. Contains ESTs, GSSs and two putative CpG islands, complete sequence.686 (0.0)
AA101348	Homo sapiens, clone IMAGE:4025624, mRNA.Homo sapiens similar to dendritic cell protein (LOC63319), mRNA. 809 (0.0)Homo sapiens dendritic cell protein (GA17), mRNA.801 (0.0)
AA127058	Homo sapiens similar to RIKEN cDNA 2610103J23 gene (LOC92140), mRNA. Homo sapiens genomic DNA, chromosome 8q23, clone: KB1907C4.535 (E-150)
AA132065	Homo sapiens chromosome 5 clone CTC-222O22, complete sequence. Homo sapiens mRNA for SMAP-5, partial cds.383 (e-103)Human DNA sequence from clone RP1-315G1 on chromosome Xq24-25. Contains a PDZ (DHR, GLGF) domain protein pseudogene, the API3 gene for apoptosis inhibitor 3 (XIAP, HILP), a putative novel gene, ESTs, STSs, GSSs and a putative CpG island, complete sequence.44 (0.15)

AA143509	Homo sapiens pyrroline-5-carboxylate synthetase (glutamate gamma-semialdehyde synthetase) (PYCS), mRNA.993 (0.0)Human DNA sequence from clone RP11-7D5 on chromosome 10, complete sequence.696 (0.0)
T53404	Homo sapiens, hypothetical protein from clone 643, clone MGC:5115 IMAGE:2984805, mRNA, complete cds.702 (0.0)
T57815	Homo sapiens similar to U5 snRNP-specific 40 kDa protein (hPrp8-binding); prp8, U5 snRNP-specific 40 kDa protein (H.sapiens) (LOC127574), mRNA.328 (3e-60)Homo sapiens U5 snRNP-specific 40 kDa protein (hPrp8-binding)(HPRP8BP), mRNA.Homo sapiens U5 snRNP-specific 40 kDa protein mRNA, complete cds.200 (6e-49)
T67053	Homo sapiens Chromosome 22q11.2 BAC Clone 142e2 In IGLC Region, complete sequence. Human lambda-immunoglobulin constant region complex (germline).775 (0.0)Homo sapiens cDNA FLJ32612 fis, clone STOMA2000088, highly similar to IG LAMBDA CHAIN C REGIONS.771 (0.0)Homo sapiens mRNA for immunoglobulin lambda-3 surrogate light chain, 3 exon form, 119 bp second exon.Homo sapiens mRNA for immunoglobulin lambda-3 surrogate light chain, 3 exon form, 122 bp second exon.Homo sapiens germline mRNA for immunoglobulin lambda-1 chain constant region, Daudi cell line.737 (0.0)
T81091	Homo sapiens coatomer protein complex, subunit alpha (COPA), mRNA.531 (5e-28)
T96829	Homo sapiens similar to cyclin-E binding protein 1 (H. sapiens) (MGC14386), mRNA.178 (3e-48)
W49619	Homo sapiens cadherin 2, type 1, N-cadherin (neuronal) (CDH2), mRNA.Homo sapiens chromosome 17, clone CTD-2023G8, complete sequence.825 (0.0)
W69906	Homo sapiens uridine monophosphate kinase (UMPK), mRNA. Homo sapiens uridine-cytidine kinase 2 (UCK2) mRNA, complete cds.373 (e-101)
W96107	Homo sapiens Sec61 gamma (SEC61G), mRNA.609 (e-172)

TABLE 5

92 Genes Identified By Single-Variable Logistic Regression

Accession Number	NAME
AA010393	Homo sapiens chromosome 17, clone hRPK.214_O_1, complete sequence. 599 (e-69)
AA019591	Homo sapiens PAC clone RP5-1186P10 from 7q11.21-q21.1, complete sequence.605 (e-170)Homo sapiens, GTF2I repeat domain-containing 1, clone MGC:9316 IMAGE:3913745, mRNA, complete cds. Homo sapiens general transcription factor 3 (GTF3) mRNA, complete cds. Homo sapiens RBAP2 (RBAP2) mRNA, complete cds. Homo sapiens putative transcription factor (WBSCR12) mRNA, complete cds. Homo sapiens muscle TFII-I repeat domain-containing protein 1 mRNA, complete cds. 414 (e-113)
AA024832	Homo sapiens mRNA; cDNA DKFZp586I0324 (from clone DKFZp586I0324). 660 (0.0)Human DNA sequence from clone RP11-23P11 on chromosome 13 Contains part of the GPC6 gene encoding Glypican 6, ESTs, STSs and GSSs, complete sequence.365 (2e-98) Homo sapiens leukocyte differentiation antigen (CD84) gene, partial cds.40 (1.4)
AA113339	Homo sapiens cDNA FLJ32537 fis, clone SMINT2000400, highly similar to Homo sapiens FRG1 mRNA.706 (0.0) Homo sapiens FSHD region gene 1 (FRG1), mRNA.601 (e-169)
AA115248	Homo sapiens BAC clone RP11-278G12 from 2, complete sequence. 862 (0.0) Homo sapiens cDNA FLJ31353 fis, clone MESAN2000264.
AA121704	Human DNA sequence from clone RP11-319I23 on chromosome 10, complete sequence. citb_10_k_1, complete sequence. 630 (e-178)
AA134595	Homo sapiens chromosome 5 clone CTD-2353F22, complete sequence. 64 (1e-7)
AA142875	Homo sapiens chromosome 10 clone RP11-176H12, complete sequence. Homo sapiens chromosome 19, cosmid R27516, complete sequence.46 (0.028)
AA157797	Homo sapiens chromosome 19, cosmid R29368, complete sequence.741 (0.0)Homo sapiens egf-like module containing, mucin-like, hormone receptor-like sequence 2 (EMR2), mRNA. Homo sapiens EGF-like module EMR2 (EMR2) mRNA, complete cds.256 (e-65)
AA165400	Homo sapiens hypothetical gene supported by AK021643 (LOC136384), 44 (0.14)Homo sapiens BAC clone RP11-832D10 from 7, complete sequence.
AA284268	Homo sapiens hypothetical gene supported by XM_072511 (LOC137861), mRNA.1096 (0.0) Human DNA sequence from clone RP11-31E23 on chromosome 1q31.3-32.1 Contains STSs and GSSs, complete sequence. Human DNA sequence from clone RP3-324N14 on chromosome 6q23.1-24.3 Contains part of protein 4.1-G mRNA, STSs and GSSs, complete sequence.40 (2.2)
AA284292	H.sapiens mRNA for beta-1,4-galactosyltransferase (EC 2.4.1.22). 813 (0.0) Homo sapiens UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 1 (B4GALT1), mRNA. 799(0.0)
AA404694	Homo sapiens PTK2 protein tyrosine kinase 2 (PTK2), mRNA. Human focal adhesion kinase (FAK) mRNA, complete cds. 747 (0.0) Homo sapiens PTK2 protein tyrosine kinase 2 (PTK2), mRNA. Homo sapiens focal adhesion kinase mRNA, complete cds. 731 (0.0)
AA406603	Human DNA sequence from clone RP11-475E11 on chromosome 1, complete sequence. Homo sapiens mRNA for KIAA0761 protein, partial cds.890 (0.0) Human DNA sequence from clone 1177E19 on chromosome 1p36.12-36.31. Contains the 3' part of the DNA-binding Zinc finger protein RIZ gene, ESTs, an STS, GSSs and a CpG island, complete sequence.44 (0.12) Homo sapiens fer-1-like 3, myoferlin (C. elegans) (FER1L3), mRNA. Homo sapiens myoferlin (MYOF) mRNA, complete cds, alternatively spliced.40 (1.9)
AA421783	Homo sapiens zinc finger protein 263 (ZNF263), mRNA. Homo sapiens mRNA for zinc finger protein FPM315, complete cds.791 (0.0)

AA424834	Homo sapiens DC1 (DC1) mRNA, complete cds. 787 (0.0)
AA429399	Homo sapiens chromosome 1 clone RP11-86H7, complete sequence.593 (e-167) Human DNA sequence from clone 283K11 on chromosome 6q23.1-24.3. Contains part of the EYA4 gene for eyes absent Drosophila) homolog 4. Contains ESTs and GSSs, complete sequence.42 (0.31)
AA431184	Human chromosome 14 DNA sequence BAC R-1078H9 of library RPCI-11 from chromosome 14 of homo sapiens (Human), complete sequence.579 (e-163)Homo sapiens chromosome 14 clone BAC257P13 map 14q31, complete sequence. 571 (e-160)Homo sapiens mucin and cadherin-like (MUCDHL), transcript variant 4, mRNA. 42 (0.38)Homo sapiens MUCDHL (MUCDHL) gene, complete cds, alternatively spliced.
AA435936	Homo sapiens Xp22 bins 169-171 BAC GSHB-383H3 (Genome Systems Human BAC Library) complete sequence.496 (e-138) Homo sapiens genomic DNA, chromosome 11q, clone:RP11-299I2, complete sequence.42 (0.34)
AA436158	Homo sapiens signaling adaptor protein DIP13alpha mRNA, complete cds. Homo sapiens, Similar to adaptor protein containing pH domain, PTB domain and leucine zipper motif, clone IMAGE:4295177, mRNA. Homo sapiens adaptor protein APPL mRNA, complete cds.872 (0.0)
AA436871	Homo sapiens syntaxin 3A (STX3A), mRNA. Homo sapiens genomic DNA, chromosome 11q, clone:CMB9-26D16, complete sequence.617 (e-174)
AA443193	Human DNA sequence from clone RP11-164H16 on chromosome 6, complete sequence.662 (0.0)
AA443285	Homo sapiens hypothetical protein FLJ10769 (FLJ10769), mRNA. 777 (0.0) Homo sapiens cDNA FLJ14198 fis, clone NT2RP3002512. 769 (0.0)
AA453607	Human DNA sequence from clone RP1-180E22 on chromosome 6p11.2-12.3. Contains the 3' part of the gene for a novel protein with possible Calmodulin like calcium-binding domains, the gene KIAA0057, ESTs, STSs, GSSs and a putative CpG island, complete sequence.375 (e-101)
AA453748	Homo sapiens 12q24.1 BAC RPCI11-946P6 (Roswell Park Cancer Institute Human BAC Library) complete sequence.777 (0.0)Homo sapiens KE03 protein mRNA, partial cds.121 (5e-25)Homo sapiens NY-REN-25 antigen (NY-REN-25), mRNA. 115 (3e-23)
AA454579	Homo sapiens genomic DNA, chromosome 11q, clone:CTD-2313N18,561 (e-157)
AA454625	Homo sapiens genomic cytochrome P450, subfamily IIIA (niphedipine oxidase) (CYP3A) on chromosome 7. Homo sapiens hypothetical protein MGC5521 (MGC5521), mRNA. Homo sapiens cDNA FLJ32683 fis, clone TESTI2000120, weakly similar to Homo sapiens mRNA for zinc finger 3 (ZF3 gene).470 (e-132)
AA455119	Homo sapiens COP9 (constitutive photomorphogenic, Arabidopsis, homolog) subunit 7A (COPS7A), mRNA.1017 (0.0)
AA457374	Homo sapiens BAC clone RP11-57B24 from 4, complete sequence.702 (0.0). Human DNA sequence from clone RP11-567B20 on chromosome 1, complete sequence. 46 (0.023)
AA459950	Homo sapiens, ribosomal protein L4, clone MGC:15542 IMAGE:3050317, mRNA, complete cds.Human DNA sequence from clone RP4-599G15 on chromosome 1p12-13.2, complete sequence.
AA460365	Homo sapiens ALS2CR4 mRNA, complete cds,.664 (0.0)Homo sapiens potential LAG1 interactor mRNA, partial cds.113 (E-22)Homo sapiens neurexin III-alpha gene, partial cds.44 (0.082)
AA463958	Human DNA sequence from clone RP4-564M11 on chromosome 1p31.1 Contains 3' end of PIGK(phosphatidylinositol glycan, class K) gene, a novel mRNA, part of a gene similar to sialyltranferase, ESTs, CA repeat, STSs and GSSs, complete sequence.345 (2e-92)
AA482325	Homo sapiens hypothetical gene supported by U18919; AL360167; BC006354; NM_025233 (LOC95925), mRNA. Homo sapiens nucleotide binding protein (NBP), mRNA. Homo sapiens nucleotide binding protein (NBP) mRNA, complete cds.680 (0.0) Human chromosome 17q12-21 mRNA, clone pOV-2, partial cds.664 (0.0)

AA488526	Homo sapiens, Similar to nucleolar phosphoprotein p130, clone MGC:5049 IMAGE:2900024, mRNA, complete cds. 741 (0.0)Human mRNA for KIAA0035 gene, partial cds.Homo sapiens similar to ORF (H. sapiens) (LOC118975), mRNA. 733 (0.0)Human DNA sequence from clone RP11-302K17 on chromosome 10, complete sequence.
AA488645	Homo sapiens transcriptional regulatory protein p54 mRNA, complete cds.1027 (0.0) homo sapiens NGFI-A binding protein 1 (EGR1 binding protein 1) (NAB1), mRNA. Homo sapiens NGFI-A binding protein 1 (EGR1 binding protein 1) (NAB1), mRNA. 1013 (0.0) Human transcriptional repressor (NAB1) NAB1 mRNA, complete cds.999 (0.0) Homo sapiens cell-line KG1 transcriptional regulatory protein p54 mRNA, complete cds.684 (0.0)
AA489246	Homo sapiens serine protease TADG15 mRNA, complete cds.1076 (0.0) Homo sapiens suppression of umorigenicity 14 (colon carcinoma, matriptase, epithin) (ST14), mRNA. Homo sapiens, Similar to uppression of tumorigenicity 14 (colon carcinoma, matriptase, epithin), clone IMAGE:2960020, mRNA, artial cds. Homo sapiens mRNA for prostamin, complete cds.1068 (0.0) Homo sapiens suppression oftumorigenicity 14 (colon carcinoma, matriptase, epithin) (ST14), mRNA. Homo sapiens matriptase mRNA, complete cds.1031 (0.0) Homo sapiens membrane-type serine protease 1 mRNA, complete cds. 1027 (0.0)Human SNC19 mRNA sequence.333 (e-88)
AA489661	Human DNA sequence from clone RP11-487I5 on chromosome 10, complete sequence.799 (0.0)
AA496780	Homo sapiens, RAB7, member RAS oncogene family, clone MGC:8453 IMAGE:2821435, mRNA, complete cds. 1005 (0.0)
AA504894	Homo sapiens, Similar to RIKEN cDNA 5830420C20 gene, clone MGC:10104 IMAGE:3898917, mRNA, complete cds.870 (0.0) Homo sapiens mannosidase, beta A, lysosomal (MANBA) gene, and ubiquitin-conjugating enzyme E2D 3 (UBE2D3) genes, complete cds.38 (6.9)
AA599093	Human DNA sequence from clone RP11-568G11 on chromosome 1, complete sequence.918 (0.0)Homo sapiens cDNA FLJ11866 fis, clone HEMBA1006973, highly similar to Homo sapiens rab3-GAP regulatory domain mRNA. 335 (2e-89)
AA609067	Homo sapiens cDNA FLJ25432 fis, clone TST06444. 779 (0.0)Homo sapiens chromosome 10 clone RP11-216P13, complete sequence.52 (4e-4)
AA609134	BAC sequence from the SPG4 candidate region at 2p21-2p22 BAC 559D11 of RPCI-11 library from chromosome 2 of Homo sapiens (Human), complete sequence.670 (0.0)Homo sapiens baculoviral IAP repeat-containing 6 (BIRC6), mRNA. Homo sapiens ubiquitin-conjugating BIR-domain enzyme APOLLON mRNA, complete cds. Homo sapiens mRNA for KIAA1289 protein, partial cds. 238 (3e-60)
AA621335	Homo sapiens hypothetical protein FLJ11088 (FLJ11088), mRNA.747 (0.0)
AA705060	Homo sapiens similar to calcium binding protein P22 (LOC115833), mRNA. Homo sapiens calcium binding protein P22 (CHP), mRNA. Human calcium-binding protein chp mRNA, complete cds.513 (e-143) Human DNA sequence from clone RP11-288H12 on chromosome 6 Contains the 3' part of the IGF2R (insulin-like growth factor 2 receptor) gene, a gene for an organic cation transporter protein, ESTs, STS, GSSs and CpG islands, complete sequence.Homo sapiens IGF2R gene, complete cds.367 (3e-99)
AA708310	Human DNA sequence from clone RP11-337C18 on chromosome 1, complete sequence. 404 (e-110)
H09747	Human DNA sequence from clone CTB-1048E9 on chromosome 22 Contains an RPS3A (Ribosomal Protein S3A) pseudogene, the gene for a novel protein similar to ASPH (aspartate beta-hydroxylase, EC 1.14.11.16), the gene for anovel protein, ortholog of mouse tuftelin-interacting protein 10 (similar to worm C07E3.1A) two more novel genes, ESTs, STSs, GSSs and three putative CpG islands, complete sequence.567 (e-159)Homo sapiens mRNA; cDNA DKFZp761P039 (from clone DKFZp761P039); partial cds.
H09818	Homo sapiens chromosome 5 clone CTB-47B11, complete sequence.343 (4e-92)Homo sapiens hypothetical protein PRO1331 (PRO1331), mRNA.321 (e-85)
H10335	Homo sapiens Chromosome 11p14.3 PAC clone pDJ292d23, complete sequence.595 (e-168) Homo sapiens RAD21 homolog (S. pombe) (RAD21), mRNA.40 (1.2)

H17335	Homo sapiens chromosome 17 map 17q21.1, complete sequence. Homo sapiens microtubule-associated protein tau (MAPT), mRNA.525 (e-146)
H23277	Homo sapiens chromosome 2 clone RP11-295N18, complete sequence.547 (e-153) Homo sapiens activin A receptor, type II (ACVR2), mRNA.Human osteosarcoma mRNA for activin typeII A receptor, complete cds.
H29292	Homo sapiens similar to ecotropic viral integration site 5 (H. sapiens) (LOC126595), mRNA.749 (0.0) Homo sapiens EVI5 homolog mRNA, complete cds.636 (e-180)
H41096	Homo sapiens chromosome 10 clone RP11-369L1, complete sequence. 714 (0.0)
H53141	Human DNA sequence from clone RP1-297M16 on chromosome 6 Contains STSs and GSSs, complete sequence.96 (3e-17)
H58736	Human chromosome 14 DNA sequence BAC C-3059L23 of library CalTech-D from chromosome 14 of Homo sapiens (Human), complete sequence.587 (e-165).Homo sapiens DMR protein mRNA, complete cds.573 (e-161)
H65834	Homo sapiens chromosome 4 clone RP11-389O4, complete sequence.50 (0.001)
H70815	Human DNA sequence from clone RP11-571F15 on chromosome 9, complete sequence.551 (e-154)Human DNA sequence from clone RP11-507C10 on chromosome 6q25.2-26, complete sequence. Human DNA sequence from clone RP11-108L7 on chromosome 10. contains part of the gene for a novel Insulin-like growth factor binding type protein with Kazal-type serine protease inhibitor domain, the gene for a novel protein similar to rat tricarboxylate carrier, the gene for a novel PDZ (DHR, GLGF) domain protein, the gene for a novel protein similar to KIAA0552, KIAA0341 and Fugu hypothetical protein 2, the gene for a novel protein similar to Plasmodium POM1 and C. elegans F46G11.1, a putative novel gene, the EMA4G gene for semaphorin 4G and a novel gene. Contains ESTs, STSs, GSSs and seven putative CpG islands, complete sequence. Human DNA sequence from clone RP1-310O13 on chromosome 20q11.2 Contains part of) four novel genes, a putative novel gene, ESTs, STSs, GSSs and four putative CpG islands, complete sequence. Homo sapiens DNA sequence from clone 394P21 on chromosome 1p36.12-36.13. Contains the PAX7 gene, locus D1S2644, ESTs and STSs, complete sequence.107 (8e-21)
H93463	Human DNA sequence from clone RP11-108F13 on chromosome 1, complete sequence.143 (e-31) Human DNA sequence from clone 272L16 on chromosome 1q32.1-32.3. Contains the 3' end of the LAMB3 gene for Laminin, Beta 3 (Nicein, Kalinin, BM600) and a novel Rat Ca2+/Calmodulin dependent Protein Kinase LIKE gene. Contains ESTs, STSs, GSSs, genomic marker D1S491 and a ca repeat polymorphism, complete sequence.125 (3e-26)
H95989	Homo sapiens similar to embryonic seven-span transmembrane protein-like protein (H. sapiens) (LOC135428), mRNA. Human DNA sequence from clone RP1-302G2 on chromosome 6p11.2-21.1, complete sequence. 474 (e-131) Human 90 kD heat shock protein gene, complete cds.184 (3e-44)
N21548	Human DNA sequence from clone 301K23 on chromosome 1p35.1-36.21. Contains the 5' part of a novel gene similar to predicted yeast and worm genes. Contains ESTs and GSSs, complete sequence.1017 (0.0)
N38891	Homo sapiens PAC clone RP4-701O16 from 7q33-q36, complete sequence.821 (0.0)
N56882	Homo sapiens BAC clone RP11-467H10 from 7, complete sequence.811 (0.0)Homo sapiens clone RP4-802G15, complete sequence.763 (0.0)
N58283	Homo sapiens BAC clone RP11-273D4 from 2, complete sequence.359 (6e-97)Homo sapiens similar to ribosomal protein S2; 40S ribosomal protein S2 (H. sapiens (LOC139909), mRNA.40 (0.69)
N66933	Homo sapiens chromosome 19, cosmid R31237, complete sequence.369 (e-99) Homo sapiens candidate tumor suppressor pp32r1 (PP32R1) gene, complete cds.135 (4e-99)
N94428	Homo sapiens E1A binding protein p300 (EP300), mRNA. 500 (e-139)
R08891	Homo sapiens KIAA0676 protein (KIAA0676), mRNA.593 (e-167)
R09585	Homo sapiens rec (LOC51201), mRNA.579 (e-163) Homo sapiens genomic DNA of 8p21.3-p22 anti-oncogene of hepatocellular colorectal and non-small cell lung cancer, segment 3/11.264 (6e-68)
R22271	Human DNA sequence from clone RP11-45A16 on chromosome 9q32-33.3, complete sequence.428 (e-117)

R28669	Human DNA sequence from clone RP5-991C6 on chromosome 6q14.1-15. Contains the gene for a novel protein similar to C. elegans F55A12.9 (Tr:P91086), an RPL10 (60S ribosomal protein L10) pseudogene, ESTs, STSs, GSSs and a putative CpG island, complete sequence.240 (6e-61)
R36449	Human DNA sequence from clone RP5-1169J3 on chromosome 11p13, complete sequence. 597 (e-168)Homo sapiens chromosome 19 clone CTC-548K16, complete sequence.62 (4e-7)Human DNA sequence from clone RP4-609E1 on chromosome 1p31.2-32.1, complete sequence. 60 (e-6)Homo sapiens BAC clone GS1-207A4 from 7p11.2-p21, complete sequence. Human beta-tubulinseudogene.58 (6e-6)
R43525	Homo sapiens kinesin heavy chain member 2 (KIF2), mRNA. 515 (e-144)H.sapiens mRNA for kinesin-2.
R44132	Homo sapiens chromosome 18, clone RP11-756M1, complete sequence.379 (e-102) Human DNA sequence from clone RP11-278E14 on chromosome 6, complete sequence.44 (0.086)
R51080	Homo sapiens hypothetical gene supported by AF086185; BC011266 (LOC93556), mRNA. 76 (3e-11)
R56219	Human DNA sequence from clone RP11-88K15 on chromosome 6, complete sequence.46 (0.028)Human DNA sequence from clone 37J18 on chromosome 1p36.2-36.3.Contains a putative novel gene, ESTs and GSSs, complete sequence.44 (0.11)
R56432	Homo sapiens BAC clone RP11-68E19 from 2, complete sequence. 40 (1.7)Human DNA sequence from clone RP11-146P20 on chromosome 10, complete sequence.Human DNA sequence from clone RP11-569H20 on chromosome X, complete sequence.Human DNA sequence from clone RP11-99J16 on chromosome 1, complete sequence.Human DNA sequence from clone RP3-377F16 on chromosome 22 Contains part of one or two novel genes, ESTs and GSSs, complete sequence.
R60053	Homo sapiens hypothetical gene supported by AF086442; AK022764; AK022851 (LOC136361), mRNA. 498 (e-138)Homo sapiens BAC clone RP11-511P7 from 7, complete sequence.Homo sapiens cDNA FLJ12789 fis, clone NT2RP2001947.
R60927	Homo sapiens mRNA full length insert cDNA clone EUROIMAGE 1603443.178 (3e-42)
R64066	Homo sapiens genomic DNA, chromosome 11q, clone:RP11-823O21, complete sequence.739 (0.0)H.sapiens CpG island DNA genomic Mse1 fragment, clone 191e9, forward read cpg191e9.ft1b. 48 (0.008)Homo sapiens membrane-bound aminopeptidase P (XNPEP2) gene, complete cds. 46 (0.030)
R92455	Homo sapiens, LIM protein (similar to rat protein kinase C-binding enigma), clone MGC:2010 IMAGE:3345715, mRNA, complete cds.408 (e-111)Human DNA sequence from clone RP11-334A14 on chromosome 1, complete sequence.74 (e-10)
R94943	Human chromosome 14 DNA sequence BAC R-841O20 of library RPCI-11 from chromosome 14 of Homo sapiens (Human), complete sequence.599 (e-169)
R98628	Homo sapiens chromosome 4 clone RP11-397E7, complete sequence.733 (0.0) Human DNA sequence from clone RP3-341D10 on chromosome X Contains a gene for a novel protein, part of the gene for a protein similar to ADP ribosylation factor 3, part of a gene similar to HTF9C and a CpG island, complete sequence. Homo sapiens BAC clone GS1-155M11 from 7q21-q22, complete sequence.174 (e-141) Homo sapiens genomic protocadherin alpha cluster (PCDHA@) on chromosome 5.167 (e-138)
R99918	Human DNA sequence from clone RP11-552I13 on chromosome 10, complete sequence. 531 (e-148)
T50370	Human DNA sequence from clone RP5-1069P2 on chromosome 20 Contains the 5' end of the STK4 gene for serine/threonine kinase 4, the gene for outer mitochondrial membrane translocase HTOM34P, the gene for a novel PABPC1 (poly(A)-binding protein, cytoplasmic 1) (PABPL1) like protein. ESTs, STSs, GSSs and a putative CpG island, complete sequence. 349 9e-93)
T55592	Homo sapiens chromosome 19 clone CTD-2192J16, complete sequence.54 (5e-4)
T61792	Homo sapiens pyruvate dehydrogenase kinase 4 mRNA, 3' untranslated region, partial sequence.603 (e-160)
T68461	Homo sapiens fibrinogen, B beta polypeptide (FGB) gene, complete cds.565 (e-159)

T71680	Homo sapiens BAC clone RP11-111H13 from 2, complete sequence.829 (0.0) Homo sapiens mitochondrial ribosomal protein L30 (MRPL30), mRNA.456 (e-126) Homo sapiens similar to testican 3 (LOC115443), mRNA. Homo sapiens testican 3 (HSAJ1454), mRNA.40 (1.8)
T86983	Homo sapiens chromosome 17, clone HCIT524C5, complete sequence.626 (e-117)Human carnitine palmitoyltransferase (CPT1) mRNA, complete cds.82 (4e-13)
T90641	Human DNA sequence from clone RP11-337C18 on chromosome 1, complete sequence.498 (e-138)Homo sapiens similar to SMHS2 (H. sapiens) (LOC121883), mRNA.40 (1.0)
T96711	Homo sapiens hypothetical protein FLJ14153 (FLJ14153), mRNA. Homo sapiens cDNA FLJ14153 fis, clone NT2RM1000092, weakly similar to MULTIDRUG RESISTANCE PROTEIN 2.676 (0.0) Homo sapiens mRNA for SMAP-4, complete cds.626 (e-177)
W31919	Homo sapiens chromosome 4 clone RP11-240A2, complete sequence.44 (0.11) Homo sapiens partial TTN gene for titin.42 (0.44)
W56308	Human chromosome 14 DNA sequence BAC R-840I19 of library RPCI-11 from chromosome 14 of Homo sapiens (Human), complete sequence.511 (e-142) Homo sapiens gastrointestinal glutathione peroxidase (GPX2) gene, complete cds. 220 (2e-55) Homo sapiens T-cell receptor alpha delta locus from bases 250472 to 501670 (section 2 of 5) of the Complete Nucleotide Sequence.40 (1.7)
W99364	Human DNA sequence from clone RP11-379P1 on chromosome 9, complete sequence. 686 (0.0)

TABLE 6

Four Groups Identified By Multidimensional Analysis

Group	Accession	Name
Group 1	R34801	Homo sapiens, clone IMAGE:4281881, mRNA. 416 (e-114)Homo sapiens, clone IMAGE:4282266, mRNA.Homo sapiens cDNA FLJ11047 fis, clone PLACE1004510, highly similar to Homo sapiens cofactor of initiator function mRNA.
Group 1	AA416585	Hmo sapiens ACE-related carboxypeptidase ACE2 mRNA, complete cds. 761 (0.0)Homo sapiens angiotensin I converting enzyme (peptidyl-dipeptidase A) 2 (ACE2), mRNA.
Group 1	W30935	Homo sapiens cDNA: FLJ21715 fis, clone COL10287, highly similar to AF071569 Homo sapiens multifunctional calcium/calmodulin-dependent protein kinase II delta2 isoform mRNA. 730 (0.0)Homo sapiens cDNA FLJ31080 fis, clone HSYRA2001615, highly similar to Sus scrofa Calcium/calmodulin-dependent protein kinase II delta 2-subunit mRNA. 704 (0.0)Homo sapiens calcium/calmodulin-dependent protein kinase (CaM kinase) II delta (CAMK2D), mRNA. 702 (0.0)
Group 1	R89104	Human DNA sequence from clone RP1-134N8 on chromosome 20p12. Contains STSs, GSSs and a CpG island, complete sequence. 52 (4e-4)
Group 1	AA056381	Homo sapiens CGI-04 protein (LOC51067), mRNA.476 (-132)Homo sapiens, clone MGC:22937 IMAGE:4843916, mRNA, complete cds.Homo sapiens cDNA FLJ13995 fis, clone Y79AA1002209, weakly similar to TYROSYL-TRNA SYNTHETASE (EC 6.1.1.1).
Group 1	R24258	Homo sapiens similar to protein kinase C, zeta (LOC113121), mRNA. 313 (3E-83)
Group 1	N74284	Homo sapiens similar to zinc finger protein (LOC90812), mRNA. 680 (0.0)
Group 1	N93470	Homo sapiens hypothetical protein FLJ10948 (FLJ10948), mRNA. 460 (E-127)
Group 1	N67578	Human chromosome 14 DNA sequence BAC R-747H7 of library RPCI-11 from chromosome 14 of Homo sapiens (Human), complete sequence. 48 (0.006)Human aquaporin-5 (AQP5) gene, exon 4 and complete cds.
Group 1	R23189	Homo sapiens, clone hRPK.11_A_1, complete sequence. 42 (0.41)Human DNA sequence from clone RP11-421P21 on chromosome 6, complete sequence.
Group 1	AA460289	Homo sapiens BAC clone RP11-182H20 from Y, complete sequence. 617 (-174) Homo sapiens testis transcript Y 7 (TTY7) mRNA, partial cds, alternatively spliced. 325 (E-86)
Group 1	N30621	Human DNA sequence from clone RP11-631F7 on chromosome 6 Contains STSs, GSSs and a CpG island, complete sequence. 894 (0.0)
Group 1	N62696	Homo sapiens mRNA full length insert cDNA clone EUROIMAGE 288936.801 (0.0)
Group 1	N74963	Homo sapiens chromosome 5 clone RP11-546B8, complete sequence. 831 (0.0) Human DNA sequence from clone RP1-211D12 on chromosome 20q12-13.2 Contains the 3' end of the STK4 gene encoding serine/threonine kinase 4, the KCNS1 gene encoding the potassium voltage-gated channel, delayed-rectifier (subfamily S, member 1), a gene similar to Elafin-like protein from mouse, a putative novel gene, a putative p53 responsive gene (PRG5) a CpG island, ESTs, STSs and GSSs, complete sequence. 46 (0.27)
Group 1	AA457253	Human DNA sequence from clone 281H8 on chromosome 6q25.1-25.3. Contains up to four novel genes, one with similarity to KIAA0323 and worm C30F12.1 and another with Ubiquitin-Like protein gene SMT3 (the latter in an intron of a novel gene). Contains ESTs, STSs, GSSs, a putative CpG island and genomic marker D6S1553, complete sequence.811 (0.0) Homo sapiens mRNA for KIAA0733 protein, partial cds.

Group 1	AA401488	Human DNA sequence from clone RP5-955L16 on chromosome 6 Contains an STS, GSSs and a CpG island, c mplete sequence. 48 (0.08)
Group 1	R08151	Homo sapiens BAC clone RP11-334F17 from 2, complete sequence. 454 (e-125) Human DNA sequence from clone 168L15 on chromosome 6q26-27 Contains part of RPS6KA2 (ribosomal protein S6 kinase, 90kD, polypeptide 2 (RSK3)), ESTs, STS, GSSs and CpG islands, complete sequence. 42 (0.036)
Group 1	W55997	Homo sapiens BAC clone RP11-112N16 from 2, complete sequence. 40 (1.5)Homo sapiens chromosome 16 clone RP11-105C20, complete sequence.
Group 2	H98694	Homo sapiens similar to PI-3-kinase-related kinase SMG-1 (H. sapiens) (LOC124368), mRNA. 751 (0.0). Homo sapiens LIP isoform of BLIP (BLIP) mRNA, complete cds.Homo sapiens lambda/iota protein kinase C-interacting protein mRNA, complete cds.Homo sapiens KIAA0421 mRNA, partial cds.Homo sapiens smg-1 mRNA for phosphatidylinositol 3-kinase-related protein kinase, complete cds. 747 (0.0)
Group 2	AA488073	Homo sapiens hypothetical gene supported by J05582; U60259; U60260; U60261; X52229; X80761; AF348143; AH001451; AH001452; NM_002456 (LOC115186), mRNA. 928 (0.0)Human polymorphic epithelial mucin (PEM) mRNA, complete cds.Human episialin variant B mRNA, 3' end.Human episialin variant A mRNA, 3' end.Human pancreatic mucin mRNA, complete cds. 858 (0.0)
Group 2	AA620674	Homo sapiens BAC clone RP11-546P22 from 2, complete sequence.78 (7e-9)Homo sapiens chromosome 10 clone RP11-267O2, complete sequence.Human DNA sequence from clone RP11-498G22 on chromosome 10, complete sequence.Human DNA sequence from clone RP11-380I20 on chromosome 9, complete sequence.
Group 2	R98047	Homo sapiens BAC clone CTB-103H13 from 7q31, complete sequence.611 (e-172)Homo sapiens full length insert cDNA clone YR42A07.
Group 2	H97496	Homo sapiens chromosome 2 clone RP11-337O8, complete sequence.420 (e-111)
Group 2	AA425543	Homo sapiens chromosome 8, clone RP11-498N9, complete sequence. 609 (e-172)
Group 3	W68127	Human DNA sequence from clone RP11-87H20 on chromosome 13, complete sequence.589 (E-165)Homo sapiens clone HAW1052 unknown mRNA.579 (E-162)
Group 3	AA420993	Homo sapiens Bardet-Biedl syndrome 4 (BBS4), mRNA.720 (0.0)
Group 3	AA010182	Homo sapiens 12 BAC RP11-593B8 (Roswell Park Cancer Institute Human BAC Library) complete sequence. 749 (0.0)
Group 3	H72878	Human chromosome 14 DNA sequence BAC R-552K16 of library RPCI-11 from chromosome 14 of Homo sapiens (Human), complete sequence.593 (E-167)
Group 3	AA406020	Homo sapiens interferon-stimulated protein, 15 kDa (ISG15), mRNA.833 (0.0)Human interferon-induced 17-kDa/15-kDa protein mRNA, complete cds. Human interferon-induced 15-Kd protein (ISG) gene, exon809 (0.0)
Group 3	H28734	Homo sapiens chromosome 4 clone RP11-372N22, complete sequence. Homo sapiens glutamate receptor, ionotropic, AMPA 2 (GRIA2), mRNA. Human glutamate receptor 2 (HBGR2) mRNA, complete cds.555 (E-155)
Group 3	AA026167	Homo sapiens chromosome 22q11 clone p143i13, complete sequence.712 (0.0)
Group 3	N80491	Homo sapiens KIAA0630 protein (KIAA0630), mRNA.355 (2E-95)
Group 3	AA455302	Homo sapiens p33ING1b (ING1) mRNA, complete cds.276 (E-71) Homo sapiens ING1 gene, exons 1a, 1b, 1c. Homo sapiens ING1 tumor suppressor, variant B (ING1) mRNA, complete cds.272 (2E-70)Homo sapiens growth inhibitory protein ING1 (ING1) gene, alternatively spliced exon 1b.56 (6E-6)

Group 3	H19111	Human DNA sequence from clone RP11-149P14 on chromosome 1, complete sequence.42 (0.39)Human DNA sequence from clone RP3-404K8 on chromosome 6p22.2-22.3 Contains PRL (prolactin) gene, STSs and GSSs, complete sequence.0 (1.5)
Group 3	R22579	Homo sapiens genomic sequence surrounding NotI site, clone NR5-CA3C.377 (E-102) Human DNA sequence from clone RP4-531H16 on chromosome 20p11.22-12. Contains the 3' end of the PCSK2 gene for proprotein convertase subtilisin/kexin type 2 (NEC2), the BFSP1 gene for beaded filament structural protein 1 (filensin), a protein 91/23 (mouse Dynein light chain, TCTEX-1 like) pseudogene, a Ubiquitin-40S Ribosomal protein S27A fusion protein pseudogene, ESTs, STSs, GSSs and a CpG island, complete Sequence.44 (0.098)
Group 3	AA460831	Human DNA sequence from PAC 102G20 on chromosome 1q24-q25.Contains ESTS, STSs and a predicted CpG island.353 (3E-91)
Group 3	AA679116	Homo sapiens genomic protocadherin beta cluster (PCDHB@) on chromosome 5. Homo sapiens protocadherin beta 5 (PCDHB5), mRNA.908 (0.0)
Group 3	AA054978	Homo sapiens similar to Similar to RIKEN cDNA 0710001C05 gene (H. sapiens) (LOC122961), mRNA.837 (0.0)Human DNA sequence from clone CTA-373H7 on chromosome 22q11.22-12.2. Contains STSs, GSSs and a CpG Island, complete sequence.502 (E-139)Homo sapiens collagen, type IV, alpha 6 (COL4A6), mRNA.42 (0.49)
Group 3	T98075	Homo sapiens chromosome 5 clone CTD-2093G19, complete sequence.400 (e-109)Human DNA sequence from clone RP4-707K17 on chromosome 20q13.1 Contains part of the PTPRT gene encoding a protein tyrosinephosphatase receptor type T, ESTs, STSs and GSSs, complete sequence.40 (1.4)
Group 3	W01031	Human DNA sequence from clone RP3-341E18 on chromosome 6p11.2-12.3. Contains the KIAA0936 gene for a MAK-related kinase and a novel alternatively spliced gene. Contains a putative CpG island, ESTs and GSSs, complete sequence.678 (0.0)
Group 3	N52195	Homo sapiens genomic DNA, chromosome 21q, section 104/105.Homo sapiens genomic DNA, chromosome 6p11.2-12.3-ter, Ter region, clone:T1136, complete sequence.278 (3e-72)
Group 3	T98796	Homo sapiens MADS box transcription enhancer factor 2, polypeptide C (myocyte enhancer factor 2C) (MEF2C), mRNA.369 (e-99)
Group 3	R34121	Homo sapiens mRNA for KIAA0720 protein, partial cds.490 (e-136)
Group 3	AA455172	Homo sapiens BAC clone RP11-92L24 from 2, complete sequence.674 (0.0) Human DNA sequence from clone RP11-360J12 on chromosome 9q21.12-21.32, complete sequence.56 (2e-5)
Group 3	T60111	Homo sapiens genomic DNA, chromosome 11q, clone:CMB9-26D16, complete sequence.230 (2e-58)Homo sapiens fatty acid binding protein 5 (psoriasis-associated) (FABP5), mRNA. 222 (5e-56)
Group 3	W80719	Homo sapiens BAC clone RP11-260K18 from 4, complete sequence.339 (5e-91)Homo sapiens excision and cross link repair protein (ERCC4) gene, complete genomic sequence.52 (2e-4)
Group 3	N30302	Homo sapiens similar to GTP-binding protein (LOC95301), mRNA.Homo sapiens, Similar to guanine nucleotide binding protein-like 1, clone IMAGE:4098352, mRNA.Homo sapiens genomic DNA, chromosome 6p21.3, HLA Class I region, section 12/20.Homo sapiens cDNA: FLJ22738 fis, clone HUV00522, highly similar to HUMHSPR Human GTP-binding protein (HSR1) mRNA.341 (3e-91)
Group 3	AA488171	Homo sapiens mRNA for KIAA1743 protein, partial cds.757 (0.0)
Group 3	W86997	Homo sapiens chromosome 5 clone CTC-332L22, complete sequence.751 (0.0)
Group 3	AA479920	Homo sapiens cDNA FLJ31951 fis, clone NT2RP7007177, weakly similar to Homo sapiens multiple membrane spanning receptor TRC8 mRNA.666 (0.0)

Group 3	AA436454	Homo sapiens genomic sequence surrounding NotI site, clone NR1-OH7C. 515 (e-144)
Group 3	AA169814	Homo sapiens sorting nexin 2 (SNX2), mRNA.581 (e-163)
Group 3	N68327	Human DNA sequence from clone RP11-177H22 on chromosome 10, complete sequence.787 (0.0)
Group 3	H96213	Homo sapiens similar to early development regulator 1; homolog of polyhomeotic 1; homolog of mouse Rae28 (H. sapiens) (LOC121482), mRNA.Homo sapiens early development regulator 1 (homolog of polyhomeotic 1) (EDR1), mRNA.Homo sapiens gene for polyhomeotic 1 homolog, partial cds, exon 15.519 (e-145)
Group 3	N64741	Homosapiens, slug (chicken homolog), zinc finger protein, clone MGC:17388 IMAGE:3911047, mRNA, omplete cds. Homo sapiens zinc finger protein SLUG (SLUG) gene, complete cds. 206 (e-50)
Group 3	AA164229	Homo sapiens BAC clone RP11-567F11 from 2, complete sequence.387 (e-105)
Group 3	AA496792	Human DNA sequence from clone RP11-447M12 on chromosome 9, complete sequence.799 (0.0)
Group 3	R25614	Homo sapiens chromosome 5 clone CTD-2299E8, complete sequence.230 (8e-58)
Group 3	AA131794	Homo sapiens genomic DNA, chromosome 11q clone:RP11-746C14,complete sequence.42 (0.22)
Group 3	H90627	Human DNA sequence from clone RP11-199O16 on chromosome 9, complete sequence.581 (e-163) Human DNA sequence from clone RP1-28F12 on chromosome 20q11.22-12 Contains part of the KIAA0823 gene, ESTs, STSs and GSSs, complete sequence.100 (2e-18)
Group 3	AA152351	Homo sapiens PAC clone RP5-978E18 from 7p21, complete sequence.517 (e-144)
Group 3	AA464708	Homo sapiens PNAS-7 mRNA, partial sequence.741 (0.0)
Group 3	H63518	Homo sapiens BAC clone RP11-270E5 from 2, complete sequence.Homo sapiens PAC clone RP5-819O4 from 7q33-q35, complete sequence.42 (0.37)Homo sapiens mannose receptor, C type 1 (MRC1), mRNA.38 (5.7)
Group 3	AA457138	Homo sapiens frizzled homolog 8 (Drosophila) (FZD8), mRNA. Homo sapiens FZD8 mRNA for seven-transmembrane receptor Frizzled-8, complete cds.753 (0.0)
Group 3	AA126673	Homo sapiens chromosome 5 clone CTD-2306M10, complete sequence.813 (0.0)
Group 3	AA432268	Homo sapiens BAC clone GS1-96J14 from 7p11.2-p21, complete sequence. 884 (0.0)
Group 3	AA446013	Homo sapiens ST5 gene for suppression of tumorigenicity 5, L27a gene for ribosomal protein L27a and KIAA0298 gene.Homo sapiens gene for ribosomal protein L27A, complete cds.624 (e-176)
Group 3	H47327	Homo sapiens transforming, acidic coiled-coil containing protein 1 (TACC1), mRNA.Homo sapiens mRNA for KIAA1103 protein, partial cds.468 (e-129)
Group 3	R76275	Homo sapiens ASCL3 gene, CEGP1 gene, C11orf14 gene, C11orf15 gene, C11orf16 gene and C11orf17 gene.159 (2e-36)
Group 3	AA609976	Homo sapiens similar to mitochondrial capsule selenoprotein (H.sapiens) (LOC127476), mRNA. Homo sapiens mitochondrial capsule selenoprotein (MCSP), mRNA.890 (0.0)
Group 3	H68988	Homo sapiens f-box and leucine-rich repeat protein 5 (FBXL5), transcript variant 2, mRNA.414 (e-113
Group 3	R72174	Homo sapiens, membrane interacting protein of RGS16, clone MGC:23190 IMAGE:4855079, mRNA, complete cds. Human Chromosome 16 BAC clone CIT987SK-A-363E6, complete equenceHuman Chromosome 16 BAC clone CIT987SK-A-363E6, complete sequence.527 (e-147)
Group 3	H99816	Homo sapiens procollagen-lysine, 2-oxoglutarate 5-dioxygenas (lysine hydroxylase) 2 (PLOD2), mRNA.529 (e-148)

Group 3	N30034	Homo sapiens clone RP11-494K17, complete sequence.42 (0.63)
Group 3	AA135135	Homo sapiens cDNA FLJ14750 fis, clone NT2RP3002948, weakly similar to RING CANAL PROTEIN.726 (0.0)
Group 3	W86387	Human DNA sequence from clone RP11-424H23 on chromosome 6, complete sequence.775 (0.0)
Group 3	AA400422	Homo sapiens PAC clone RP5-978E18 from 7p21, complete sequence.650 (0.0)
Group 3	AA609608	Human DNA sequence from clone RP1-292B18 on chromosome 6q24.3-25.3. Contains a 60S ribosomal protein L32 (RPL32) pseudogene, the 3' end of the gene for a novel protein similar to NADP+ dependent methylenetetrahydrofolate dehydrogenase, and formyltetrahydrofolate synthethase (MTHFD1),ESTs, STSs and GSSs, complete sequence.398 (e-108)Homo sapiens similar to SNAG1 (H. sapiens) (LOC129365), mRNA.339 (2e-90)Human DNA sequence from clone RP11-251O17 on chromosome 9 Contains 2 calponin 2 (CNN2) pseudogenes, a gene for a novel protein similar to aquaporin 7 (AQP7), part of gene for a novel protein similar to methylenetetrahydrofolate dehydrogenase(NADP+ dependent), methenyltetrahydrofolate cyclohydrolase, formyltetrahydrofolate synthetase (MTHFD1) and a CpG island, complete sequence.262 (3e-67)
Group 3	T89372	Homo sapiens protein phosphatase 2 (formerly 2A), regulatory subunit B" (PR 72), alpha isoform and (PR 130), beta isoform(PPP2R3), mRNA. Homo sapiens protein phosphatase 2A 72 kDa regulatory subunit mRNA,complete cds.630 (e-178)
Group 3	H53038	Human DNA sequence from clone RP11-501I19 on chromosome 6, complete sequence. Human DNA sequence from clone RP3-416F21 on chromosome 6q26-27. Contains the 3' end of the PDE10A gene for MTHFD1, the 5' end of a novel gene, ESTs, STSs and GSSs, complete STSs and GSSs, complete sequence.46 (0.18)
Group 3	R68381	Human DNA sequence from clone RP11-179F17 on chromosome 13, complete sequence.428 (e-117)
Group 3	N51987	Homo sapiens chromosome 10 clone RP11-397P14, complete sequence.46 (0.031)
Goup 4	T41203	Homo sapiens chromosome 4 clone RP11-678H22, complete sequence.474 (E-131)
Goup 4	AA460369	Homo sapiens, clone MGC:19524 IMAGE:4329693, mRNA, complete cds.682 (0.0)
Goup 4	H90848	Human DNA sequence from clone RP11-286F20 on chromosome 1 Contains part of a novel gene for KIAA1383 protein, ESTs, STSs and GSSs, complete sequence.82 (5e-13)
Goup 4	AA663435	Homo sapiens tripartite motif-containing 28 (TRIM28), mRNA. Homo sapiens, Similar to KRAB-ssociated protein 1, clone IMAGE:3831930, mRNA, partial cds. Homo sapiens, KRAB-associated protein 1, clone MGC:3849 IMAGE:2906325, mRNA, complete cds. 977 (0.0)Human nuclear corepressor KAP-1 (KAP-1) mRNA, complete cds. 973 (0.0)Human transcriptional corepressor hKAP1/TIF1B mRNA, complete cds. H.sapiens mRNA for TIF1beta zinc finger protein. 957 (0.0)
Goup 4	R89581	Homo sapiens BAC clone RP11-795L3 from 2, complete sequence.98 (8e-18)
Goup 4	H73591	Homo sapiens similar to cytochrome b5 outer mitochondrial membrane precursor (H. sapiens) (LOC124229), mRNA. Homo sapiens cytochrome b5 outer mitochondrial membrane precursor (CYB5-M), mRNA.341 (3e-91)
Goup 4	H48105	Homo sapiens PAC clone RP4-593H12 from 7p31, complete sequence. 383 (e-104) Homo sapiens protein similar to E.coli yhdg and R. capsulatus nifR3 (PP35), mRNA. 379 (e-103) Human PP35 mRNA, complete cds. 331 (2e-88)
Goup 4	H56152	Homo sapiens BAC clone RP11-575M4 from 7, complete sequence.432 (e-118

Goup 4	R98344	Homo sapiens, transmembrane 4 superfamily member 9, clone MGC:9300 IMAGE:3895933, mRNA, complete cds. Homo sapiens tetraspan NET-4 mRNA, complete cds.291 (3e-76) Homo sapiens tetraspan 5 (TSPAN-5), mRNA. Homo sapiens tetraspan TM4SF (TSPAN-5) gene, complete cds. 274 (6e-71)
Goup 4	AA481480	Homo sapiens, KIAA0255 gene product, clone IMAGE:3507918, mRNA. Human DNA sequence from clone RP5-836N17 on chromosome 20q11.1-11.21 Contains part of the HCK (hemopoietic cell kinase) gene, the KIAA0255 gene, a ribosomal protein L30 pseudogene, ESTs, STSs, GSSs and CpG Islands, complete sequence.954 (0.0)
Goup 4	R02069	Homo sapiens hnRNP 2H9E mRNA, complete cds.432 (e-119) Homo sapiens, Similar to heterogeneous nuclear ribonucleoprotein H3 (2H9), clone IMAGE:3927179, mRNA, partial cds.167 (8e-39)
Goup 4	AA971406	Homo sapiens hypothetical gene supported by D87682 (LOC136725), mRNA. Homo sapiens clone UWGC:djs1 or RP11-16G1 from 7p14-15, complete sequence. Human mRNA for KIAA0241 gene, partial cds. 753 (0.0)
Goup 4	R00884	Homo sapiens dihydrofolate reductase (DHFR), mRNA. Human dihydrofolate reductase gene, exon 6 and 3' flank.402 (e-110) Homo sapiens genomic DNA, chromosome 6p21.3, HLA Class I region, section 6/20. Homo sapiens genomic DNA, 237 kb segment from 6p21.3 region including HLA genes, complete sequence. 252 (e-64)
Goup 4	T61475	Homo sapiens clone 23664 and 23905 mRNA sequence. Homo sapiens cDNA: FLJ23599 fis, clone LNG15473, highly similar to AF035315 Homo sapiens clone 23664 and 23905 mRNA sequence.591 (e-166)
Goup 4	H05635	Homo sapiens chromosome 16 clone RP11-137H10, complete sequence.589 (e-166)
Goup 4	AA453603	Homo sapiens orphan nuclear receptors (NR1I2) gene, complete cds, alternatively spliced.680 (0.0) Human DNA sequence from clone RP11-90H3 on chromosome 1, complete sequence.48 (0.006)
Goup 4	N77198	Homo sapiens chromosome 17, clone RP11-272E10, complete sequence.470 (e-130) Homo sapiens chromosome 19, BAC 273239 (CIT-B-320G13), complete sequence.153 (2e-34)
Goup 4	AA035730	Homo sapiens chromosome 5 clone CTC-202F10, complete sequence.829 (0.0)
Goup 4	T86959	Homo sapiens chromosome 5 clone CTC-529P8, complete sequence.872 (0.0) Homo sapiens PAC clone RP4-547C10 from 7p21-p22, complete sequence.205 (5e-50) Homo sapiens chromosome 15 clone CTD-2306A12 map 15q21.1, complete sequence. 196 (e-47)
Goup 4	AA625806	Homo sapiens ninjurin 1 (NINJ1), mRNA. Human adhesion molecule ninjurin mRNA, complete cds.454 (e-125)
Goup 4	AA489314	Homo sapiens similar to putative (LOC96645), mRNA. Homo sapiens p25 mRNA, complete cds. Homo sapiens, Similar to gp25L2 protein, clone MGC:2142 IMAGE:2967520, mRNA, complete cds.676 (0.0) Homo sapiens sulfotransferase family, cytosolic, 1C, member 2 (SULT1C2), mRNA. Homo sapiens SULT1C sulfotransferase (SULT1C) mRNA, complete cds. 476 (e-132)
Goup 4	AA432108	Homo sapiens serine racemase (SRR), mRNA. Homo sapiens cDNA FLJ13107 fis, clone NT2RP3002501, weakly similar to THREONINE DEHYDRATASE CATABOLIC (EC 4.2.1.16).979 (0.0) Homo sapiens chromosome 17 sequence from PAC RPCI-5 1037N22 map 17q13.3 region D17S695-D17S654, complete sequence. 858 (0.0) Homo sapiens mRNA for KIAA1401 protein, partial cds.848 (0.0)
Goup 4	AA088438	Homo sapiens chromosome 16 clone CTC-508F8, complete sequence.145 (4e-32)

TABLE 7A

Association Between The Expression of The Identified Markers With Tumor Stage And Grade

BIOMARKER	*STAGE*		*GRADE*	
	NUMBER OF *CASES*	*P-VALUE*	NUMBER OF *CASES*	*P-VALUE*
p21	150	<0.0005	147	<0.0005
CYCLIN E	155	NS	153	NS
CYTOKERATIN 20	155	0.005	153	<0.005
NEUROPILIN-2	157	<0.0005	154	<0.0005
NINJURIN	147	NS	144	NS
p33ING1	146	<0.0005	145	<0.0005

TABLE 7B

Association Between The Expression of The Identified Markers With The Expression of p53 And pRB

BIOMARKER	*P53*			*UNDERPHOSPHORYLATED RB*		
	NUMBER OF CASES	KENDALL'S TAU-b	*P*	NUMBER OF CASES	KENDALL'S TAU-b	*P*
p21	*NS*			*146*	*0.316*	*<0.0005**
CYCLIN E		*0.203*	*0.001*	*NS*		
CYTOKERATIN 20	*NS*			*151*	*0.233*	*0.0005*
NEUROPILIN-2	*NS*			*152*	*0.217*	*<0.0005*
NINJURIN	*NS*			*NS*		
p33ING1	*NS*			*143*	*0.370*	*<0.0005**

*Significant also for total RB.

Data from this example revealed that expression profiling segregated superficial and invasive tumors. Moreover, these clusters rendered predictive information. Multidimensional analyses supported such clustering while also identifying those superficial tumors with expression profiles similar to invasive lesions. The analytical approach undertaken identified genes that were associated with tumor stage and grade, as well as altered p53 and pRB expression. Furthermore, p33ING1 showed a significant association with patient survival when validated in tissue microarrays. Overall, both clusters and individual targets showed clinical value for subtype classification and prognosis of patients with bladder cancer.

Following the segregation of tumor subtypes by gene clustering, analysis of data focused on gene identification methods and validation using tissue microarrays. Different algorithms were applied for gene identification providing distinct ranked gene lists. The Mann-Whitney-Wilcoxon test is a standard means for gene identification between two groups. In this case, p21 and cyclin E were selected for further validation. p21 was found to be associated with tumor stage and grade, in accordance with previous series.

Single-variable logistic regression is a standard classification/discrimination model to rank gene by their classification performance. np2 and cytokeratin-20, known soluble proteins, were selected for further validation, and their expression shown to be associated with tumor stage and grade. np-2 is a transmembrane receptors for semaphorins (mediators of neuronal guidance), and for several angiogenic factors, including vascular endothelial growth factor 145 (VEGF 145) and VEGF 165. It has been reported that osteosarcomas overexpressing np-2 had increased vascularity and poorer prognosis, suggesting that np-2 acts as a VEGF-amplifier in these tumors. An association between np-2 expression and tumor progression has also been reported for certain neoplasms, including prostate and lung cancer. The association of cytokeratin 20 with stage and grade had previously been reported for bladder cancer both in tissue and urine specimens.

Ninjurin and p33ING1 were selected among the target genes differentially expressed in the poor prognostic groups identified by multidimensional analysis. Ninjurin a nerve injury induced protein involved in neuronal growth, is

known to be altered in hepatocellular carcinoma, acute lymphoblastic leukemia, and was reported to be down-regulated by p53. Nevertheless, ninjurin was not found to be significantly associated with tumor stage and grade in the cohort of bladder cancer patients analyzed. The expression of p33ING1 was significantly associated with tumor stage and grade, as well as with overall patient survival. p33ING1 has been reported to cooperate with p53 in blocking cell proliferation and enhancing apoptosis, and it has been ascribed as candidate tumor suppressor gene. Its expression has been reported to be involved in the progression of lymphoid tumors. Down-regulation of p33ING1 was associated with pRB, p21 and cyclin E expression, but not with p53.

The SVM algorithm revealed that WNT and mitotic spindle checkpoint are among the most important pathways altered during bladder cancer progression. Deregulation of WNT was observed in the present analysis. Furthermore, WNT signaling and the mitotic spindle checkpoint are related. WNT signaling induces alignment and orientation of the mitotic spindle presumably by directly targeting the cytoskeleton. Relationships between p53 and spindle checkpoint are also being established. For example, it has been described that 53BP1, a direct p53 binding protein that contains two BRCT domains and implicated in early response to DNA damage, significantly co-localizes with CENP-E to kinetochores. 53BP1 is loaded to kinetochores in prophase, before CENP-E, and is released by mid-anaphase.

The following references were cited herein:

Cheung et al., Making and reading microarrays. Nat. Genet. 21(Suppl):15-19 (1999).

Cristiani & Shawe-Taylor, Support Vector Machines and other kernel-based learning methods, Cambridge University Press, Cambridge, (2000).

Dawson-Saunders and Trapp, Basic & Clinical Biostatistics, 2nd edition, Norwalk, Connecticut, Appleton & Lange (1994).

De las Rivas et al., Genome Research 12:567-583 (2002).

Dudoit, Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. U.C. Berkeley Dept of Statistics Technical Report #578 (2000).

Eisen et al., Proc. Natl. Acad. Sci. USA 95:14863-14868 (1998).

Felsenstein, Evolution 39:783-791 (1985).

Fleming and Lin, Biometrics 56:971-983 (2000).

Hoos et al., Am. J. Pathol. 158:1245-1251 (2001).

Kaplan and Meier, J. Am. Stat. 53:457-481 (1958).

Li & Yang, How many genes are needed for a discriminant microarray data analysis, in Methods of Microarray Data Analysis, Eds Lin et al. (Kluwer Academic), pp.137-150 (2002).

McShane et al., Clin. Cancer Res. 6:1854-1864 (2000).

Reyment & Joreskog, Applied Factor Analysis in the Natural Sciences. Cambridge University Press, Cambridge, (1996).

Saitou & Nei, Mol. Biol. Evol. 4:406-425 (1987).

Stears et al., Physiol. Genomics 3:93-99 (2000).

Tseng et al., Nucleic Acids Res. 29:2549-2557 (2001).

Tudor and Koch, Stat. Methods Med. Res. 3:345-381 (1994).

Wikman et al., Clin. Chem. 46:1555-1561 (2000).

Xie et al., Genome Res. 12:785-94 (2002).

Xiong et al., Mol. Genet. Metab. 73:239-47 (2001).

Any patents or publications mentioned in this specification are indicative of the levels of those skilled in the art to which the invention pertains. Further, these patents and publications are incorporated by reference herein to the same extent as if each individual publication was specifically and individually indicated to be incorporated by reference.

WHAT IS CLAIMED IS:

1. A method of diagnosing squamous metaplasia of bladder cancer in an individual, comprising the steps of:

collecting biological samples from said individual; and

determining the expression of caveolin-1 or keratin 10 in said samples, wherein the expression of caveolin-1 or keratin 10 in said samples indicates the presence of squamous metaplasia of bladder cancer.

2. The method of claim 1, wherein expression of caveolin-1 or keratin 10 is determined at protein level.

3. The method of claim 1, wherein expression of caveolin-1 or keratin 10 is determined at nucleic acid level.

4. A method of diagnosing bladder cancer in an individual, comprising the steps of:

collecting biological samples from said individual; and

determining in said samples the level of gene expression for a protein selected from the group consisting of zyxin, E-cadherin, moesin, cytokeratin 20, neuropilin 2, p21 and p33ING1, wherein said level of expression correlates with the stage and grade of bladder cancer in said individual.

5. The method of claim 4, wherein said gene expression is determined at protein level.

6. The method of claim 4, wherein said gene expression is determined at nucleic acid level.

7. A method of discriminating between superficial and invasive bladder cancer in an individual, comprising the steps of:

collecting biological samples from said individual; and

determining in said samples the level of gene expression for a protein selected from the group consisting of zyxin, E-cadherin, moesin, p21, cytokeratin 20, neuropilin-2 and p33ING1, wherein said protein is differentially expressed in superficial and invasive bladder cancer.

8. The method of claim 7, wherein said gene expression is determined at protein level.

9. The method of claim 5, wherein said gene expression is determined at nucleic acid level.

10. A method of predicting survival outcome of an individual having bladder cancer, comprising the steps of:

collecting biological samples from said individual; and

determining in said samples the level of gene expression for p33ING1 or moesin, wherein said level of expression correlates with survival outcome of said individual.

11. The method of claim 7, wherein said gene expression is determined at protein or nucleic acid level.

12. A method of discriminating between superficial and invasive bladder cancer in an individual, comprising the steps of:

collecting biological samples from said individual; and

determining in said samples the expression of a gene identified by an accession number selected from the group consisting of AA011414, AA021434, AA021464, AA028884, AA034115, AA035095, AA043806, AA074666, AA083385, AA101348, AA127058, AA132065, AA143509, AA147928, AA156863, AA165403, AA172210, AA190401, AA256462, AA279188, AA394148, AA402766, AA421518, AA424578, AA425861, AA430520, AA434068, AA446453, AA447696, AA449831, AA450227,

AA450265, AA454566, AA454862, AA455150, AA455281, AA456136, AA457092, AA457162, AA457725, AA458661, AA459663, AA464152, AA464192, AA465031, AA465378, AA465593, AA478268, AA485052, AA486313, AA486374, AA486761, AA487020, AA487223, AA487265, AA487899, AA489400, AA490047, AA490390, AA496359, AA496784, AA496948, AA504128, AA504617, AA598759, AA598815, AA620479, AA625981, AA629584, AA633757, AA669341, AA680322, AA682613, AA683085, AA705886, AA775415, AA862434, AA934762, AA935560, AI017703, H05769, H17158, H20652, H21040, H23366, H23880, H54093, H73731, H84444, H93463, H94897, H99502, N30811, N54338, N69283, N73536, N91962, R16165, R22439, R24543, R25377, R27552, R49144, R53889, R55763, R69307, R76314, R78514, T53404, T57815, T67053, T81091, T96829, W49619, W69906, W96107, AA010393, AA019591, AA024832, AA113339, AA115248, AA121704, AA134595, AA142875, AA157797, AA165400, AA284268, AA284292, AA404694, AA406603, AA421783, AA424834, AA429399, AA431184, AA435936, AA436158, AA436871, AA443193, AA443285, AA453607, AA453748, AA454579, AA454625, AA455119, AA457374, AA459950, AA460365, AA463958, AA482325, AA488526, AA488645, AA489246, AA489661, AA496780, AA504894, AA599093, AA609067, AA609134, AA621335, AA705060, AA708310, H09747, H09818, H10335, H17335, H23277, H29292, H41096, H53141, H58736, H65834, H70815, H93463, H95989, N21548, N38891, N56882, N58283, N66933, N94428, R08891, R09585, R22271, R28669, R36449, R43525, R44132, R51080, R56219, R56432, R60053, R60927, R64066, R92455, R94943, R98628, R99918, T50370, T55592, T61792, T68461, T71680, T86983, T90641, T96711, W31919, W56308, W99364, wherein said gene is differentially expressed at the mRNA level in superficial and invasive bladder cancer.

13. A method for identifying the presence or absence of a squamous metaplasia of bladder cancer phenotype in a cell or cells, comprising determining the expression level of caveolin-1 or keratin 10 in said cell or cells, wherein a detectable

expression level of caveolin-1 or keratin 10 in said cell or cells indicates the presence of squamous metaplasia of bladder cancer phenotype and an undetectable level of caveolin-1 or keratin 10 in said cell or cells indicates the absence of squamous metaplasia of bladder cancer phenotype.

14. A method of identifying the presence or absence of a squamous metaplasia of bladder cancer in an individual, comprising the steps of:

(a) collecting a biological sample from said individual; and

(b) determining the expression level of caveolin-1 or keratin 10 in said sample,

wherein a detectable expression level of caveolin-1 or keratin 10 in said sample indicates the presence of said squamous metaplasia of bladder cancer and an undetectable level of caveolin-1 or keratin 10 in said sample indicates the absence of said squamous metaplasia of bladder cancer.

15. The method of claim 1 or 2, wherein the expression level is a protein expression level.

16. The method of claim 1 or 2, wherein the expression level is a nucleic acid expression level.

17. A method of identifying the presence or absence of a bladder cancer in an individual, comprising the steps of:

collecting a biological sample from said individual; and

determining in said sample the level of expression of a protein selected from the group consisting of zyxin, E-cadherin, moesin, cytokeratin 20, neuropilin 2, p21 and p33ING1, wherein said level of expression indicates the presence or absence of a bladder cancer in said individual.

18. The method of claim 17, further comprising correlating said expression level with the stage and grade of bladder cancer in said individual.

19. The method of claim 17 or 18, wherein expression of said protein is determined at protein level.

20. The method of claim 17 or 18, wherein expression of said protein is determined at nucleic acid level.

21. A method of discriminating between a superficial and an invasive bladder cancer in an individual, comprising the steps of:

collecting a biological sample from said individual; and

determining in said sample the level of expression of a protein selected from the group consisting of zyxin, E-cadherin, moesin and p33ING1, wherein said protein is differentially expressed in superficial and invasive bladder cancer.

22. The method of claim 21, wherein said level of expression is a level of expressed protein.

23. The method of claim 21, wherein said level of expression is a level of nucleic acid expression.

24. A method of predicting survival outcome of an individual having bladder cancer, comprising the steps of:

collecting biological samples from said individual; and

determining in said samples the level of expression of p33ING1, wherein said level of expression correlates with survival outcome of said individual.

25. The method of claim 24, wherein expression of said protein is determined at protein or nucleic acid level.

26. A kit for identifying the presence or absence of a squamous metaplasia of bladder cancer phenotype in a cell or cells, comprising a reagent or reagents capable of determining the expression level of caveolin-1 or keratin 10 in said cell or cells, wherein detectable expression levels of caveolin-1 or keratin 10 in said samples indicates the presence of squamous metaplasia of bladder cancer phenotype and undetectable levels of caveolin-1 or keratin 10 in said cell or cells indicates the absence of squamous metaplasia of bladder cancer phenotype.

27. A kit for identifying the presence or absence of bladder cancer in an individual, the kit comprising a reagent or reagents capable of determining the level of expression of a protein selected from the group consisting of zyxin, E-cadherin, moesin, cytokeratin 20, neuropilin 2, p21 and p33ING1.

27. A kit for discriminating between superficial and invasive bladder cancer in an individual, the kit comprising a reagent or reagents capable of determining the level of expression of a protein selected from the group consisting of zyxin, E-cadherin, moesin and p33ING1.

28. A kit for predicting survival outcome of an individual having bladder cancer, the kit comprising a reagent or reagents capable of determining in said samples the level of expression of p33ING1.

29. The kit according to any one of claims 26-28 wherein at least one reagent is an antibody.

30. The kit according to any one of claims 26-28 wherein at least one reagent is a nucleic acid.

31. The kit according to any one of claims 26-28 further comprising instructions for correlating said level of expression with a clinical diagnosis.

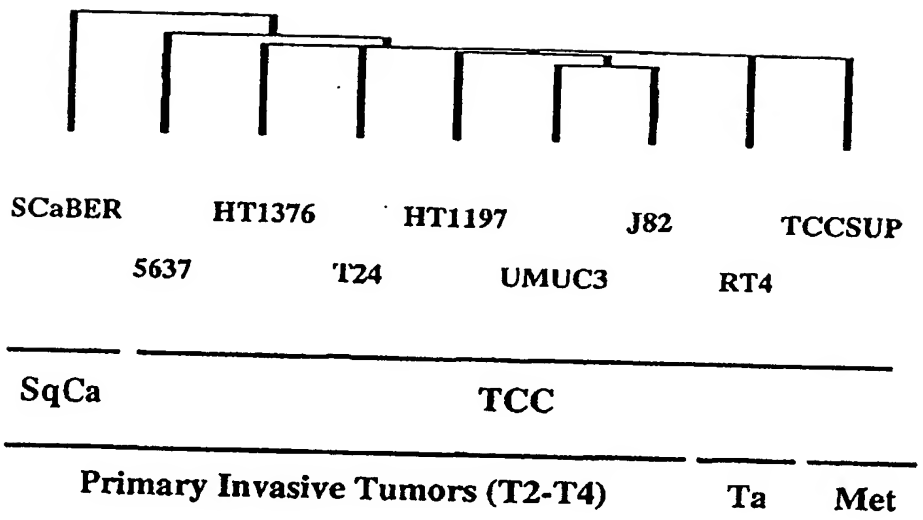

SCaBER
HT1376
HT1197
J82
TCCSUP
5637
T24
UMUC3
RT4
SqCa
TCC
Primary Invasive Tumors (T2-T4)
Ta
Met

Fig. 1

Fig. 2A

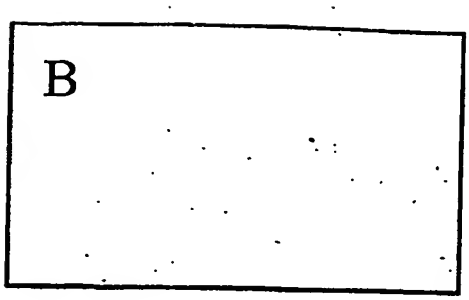


Fig. 2B

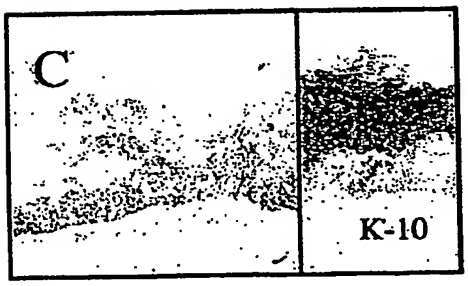


Fig. 2C

Fig. 2D

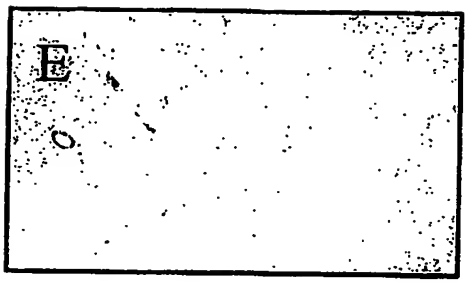


Fig. 2E

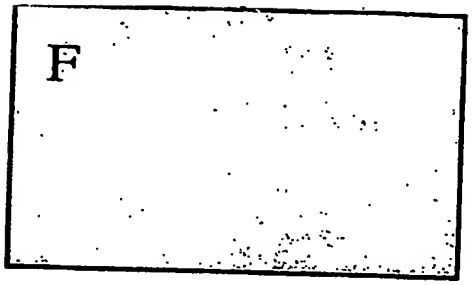
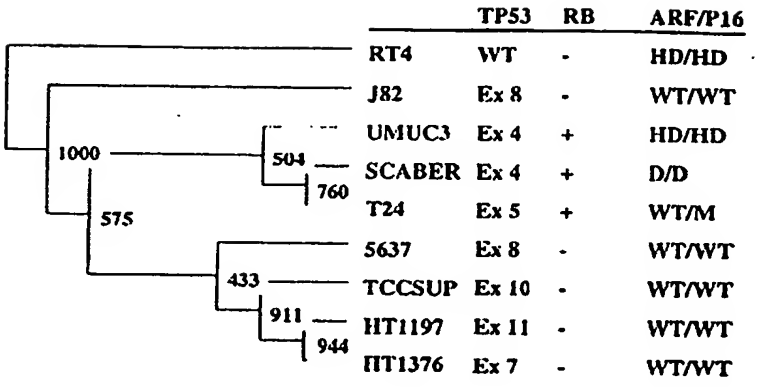


Fig. 2F

Fig. 2G

Fig. 2H



Gene Name	RT4	UMUC3	J82	T24	SCABER	5637	TCCSUP	HT1376	HT1197
MOESIN	-0.212	2.449	0.001	0.564	-0.288	-0.428	-0.83	-1.233	-1.47
PROTOCADHERIN 13	-0.02	-1.493	-1.32	-1.401	-1.47	-0.831	-0.76	-0.762	-0.875
ZYXIN	0.13	-1.295	-1.103	-1.276	-1.031	-0.733	-0.549	-0.968	-0.758

Fig. 3

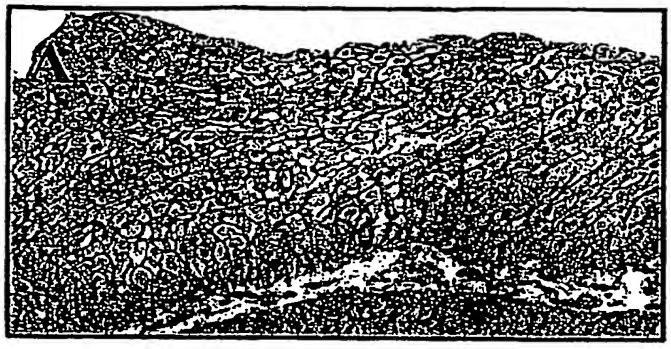


Fig. 4A

Fig. 4B

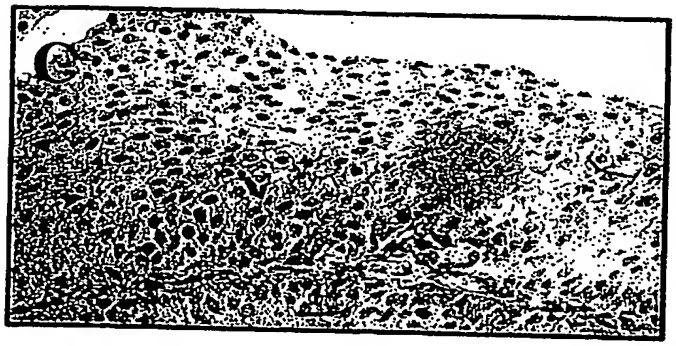


Fig. 4C

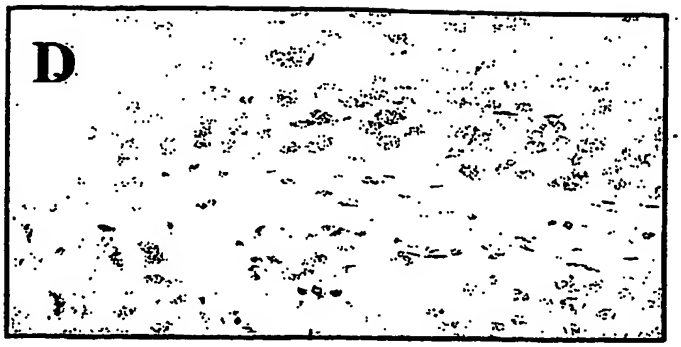


Fig. 4D

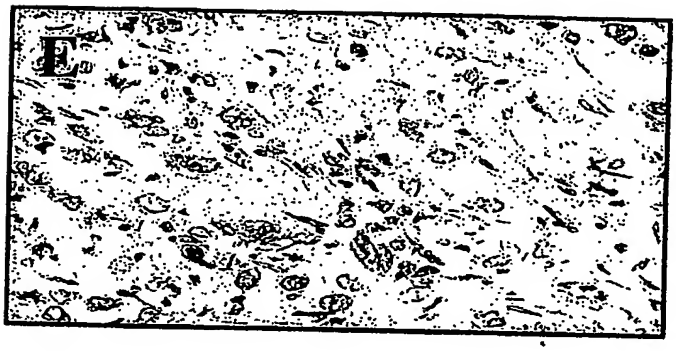
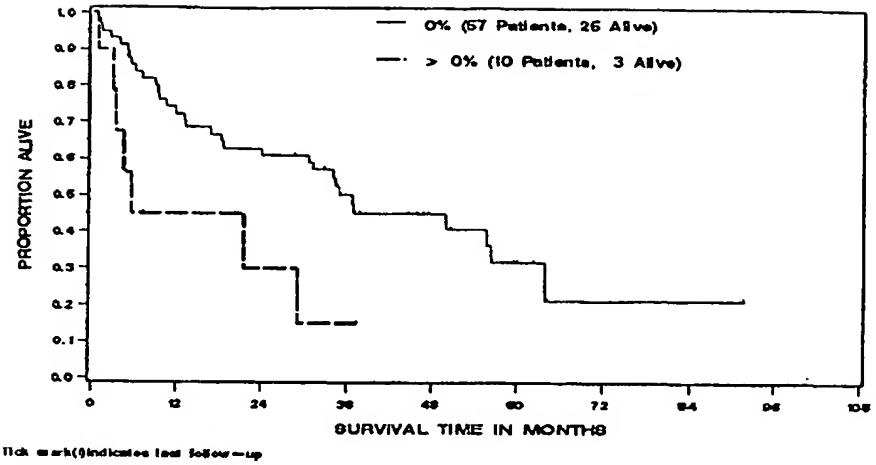


Fig. 4E

Fig. 4F



Tick mark(|)indicates last follow-up

Marker	# Events	# Censored	Median Survival (months)	p-value
Moesin (membrane)				
Undetectable	32	25	36.6	0.01
Detectable (>0%)	7	3	5.5	

Fig. 5

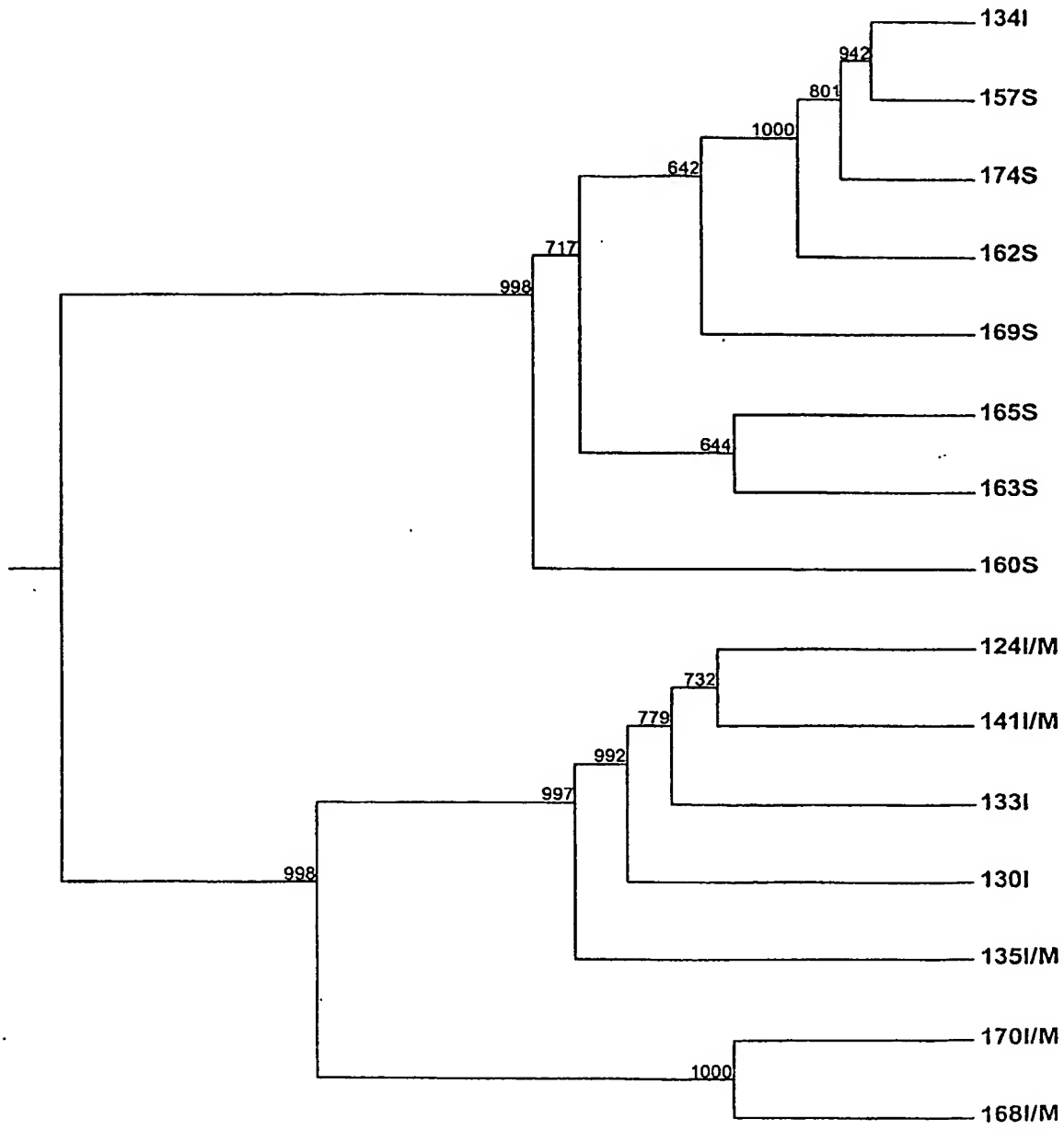

134I
942
801
157S
1000
642
174S
717
162S
998
169S
165S
644
163S
160S
124I/M
732
141I/M
779
992
133I
997
998
130I
135I/M
170I/M
1000
168I/M

Fig. 6A

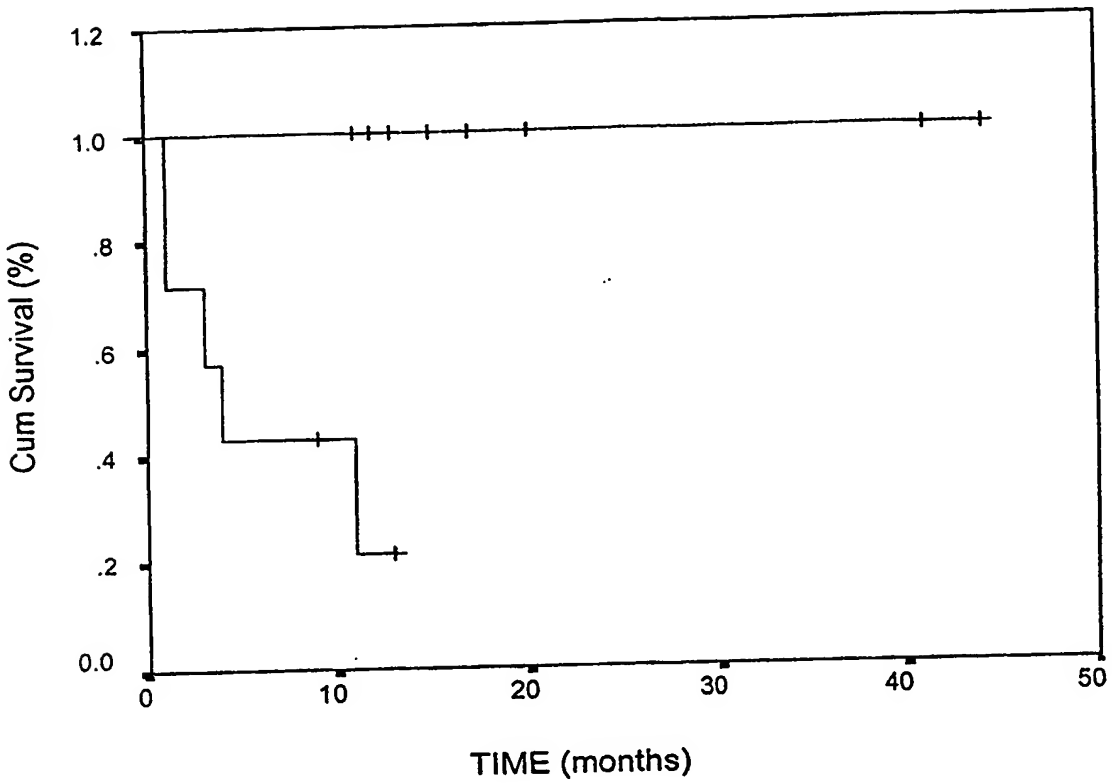
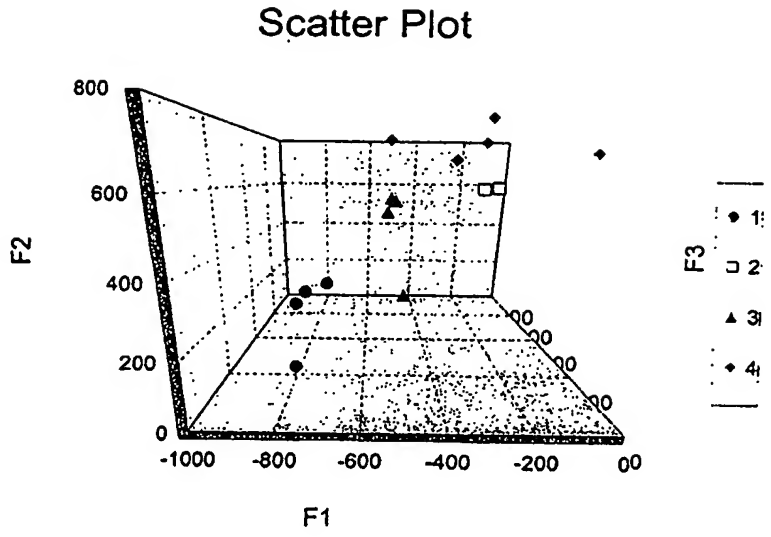

1.2
1.0
.8
.6
.4
.2
0.0
Cum Survival (%)
0
10
20
30
40
50
TIME (months)
Bootstrap clusters
2
censored
1
censored

Fig. 6B


Scatter Plot
800
600
400
200
0
F2
-1000
-800
-600
-400
-200
F1
F3

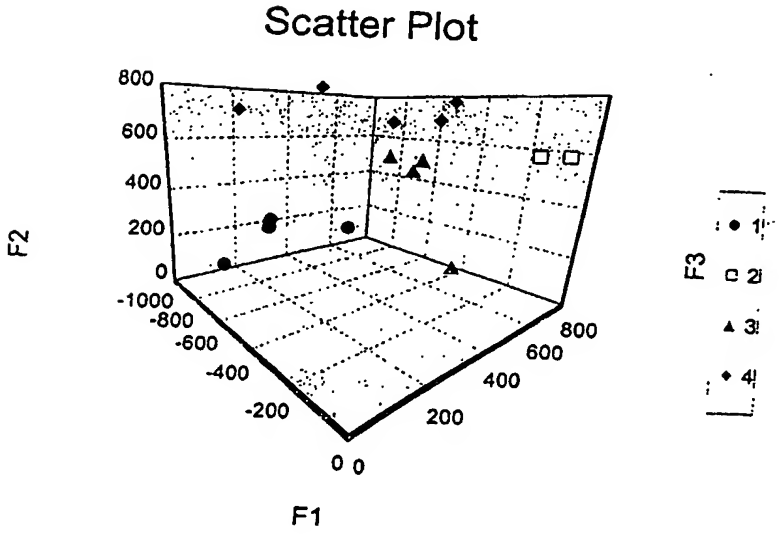

Scatter Plot
800
600
400
200
0
F2
-1000
-800
-600
-400
-200
0
F1
F3

Fig. 6C

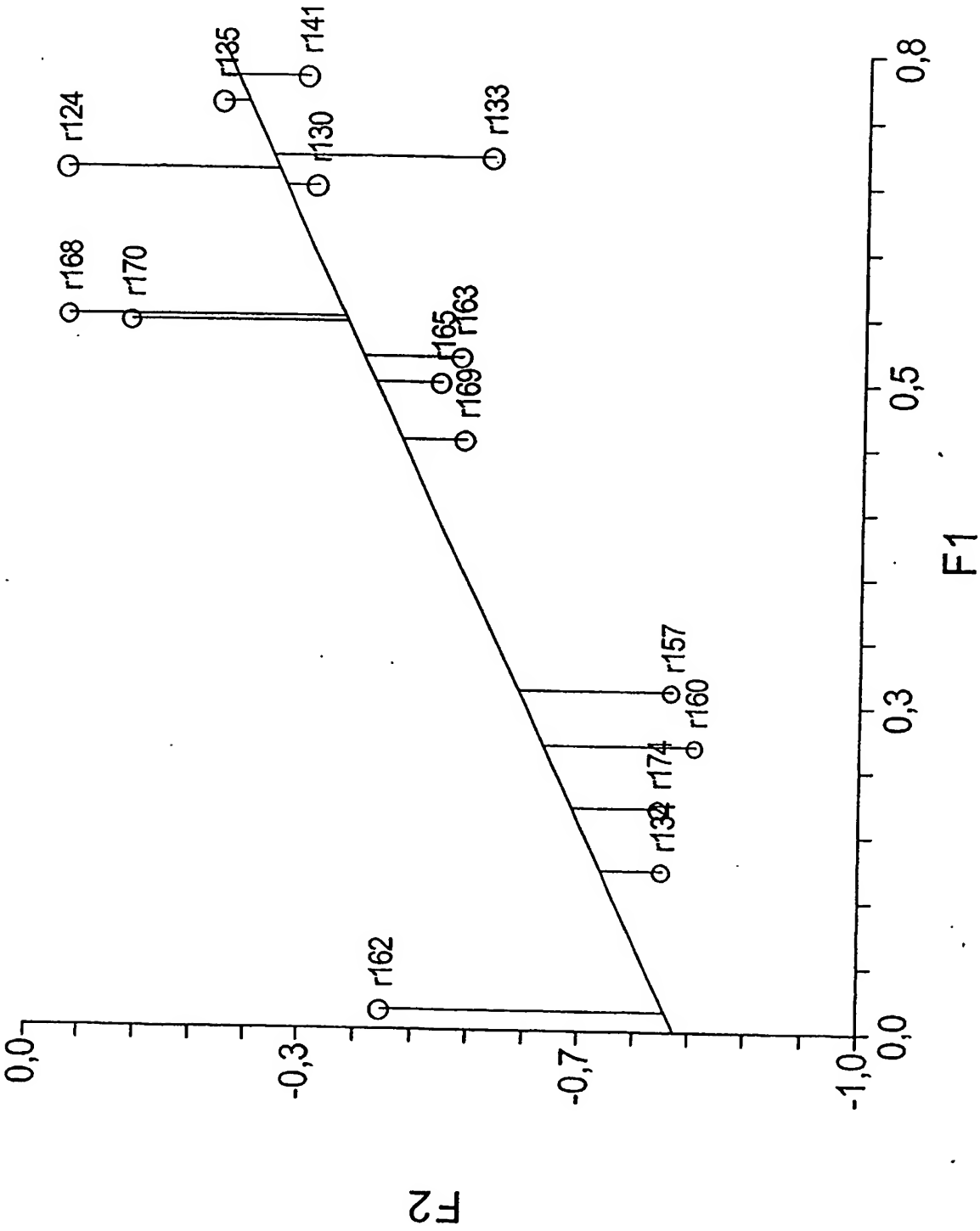

r124
r135
r141
r133
r130
r168
r170
r165
r163
r169
r157
r160
r174
r134
r162
0,0
-0,3
-0,7
-1,0
0,0
0,3
0,5
0,8
F1
F2

Fig. 6D

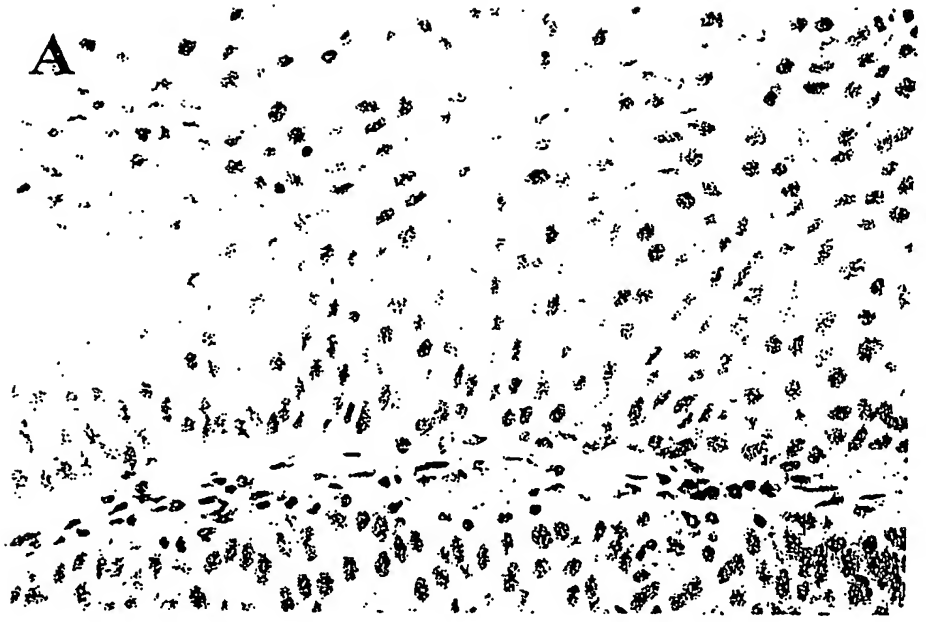


Fig. 7A

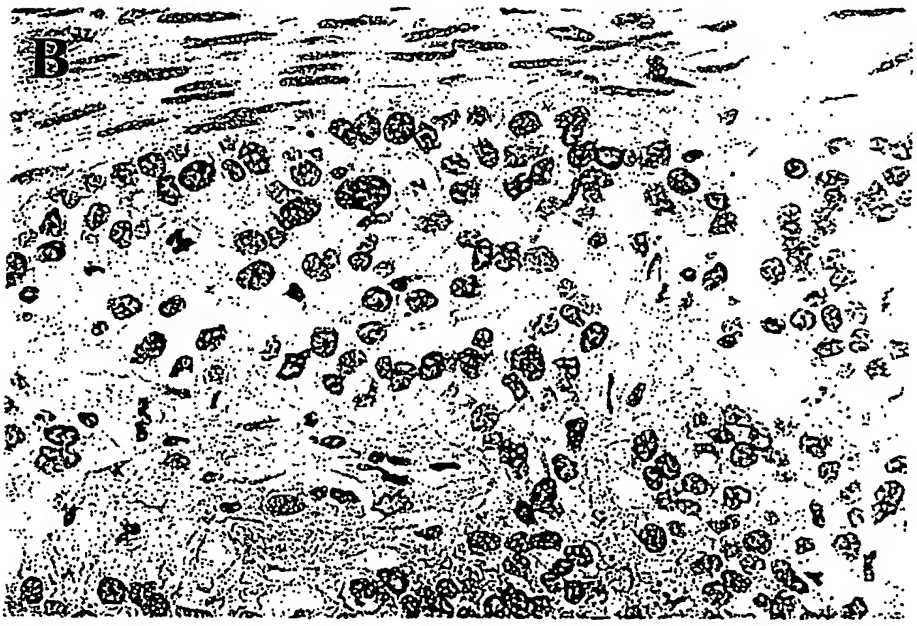


Fig. 7B

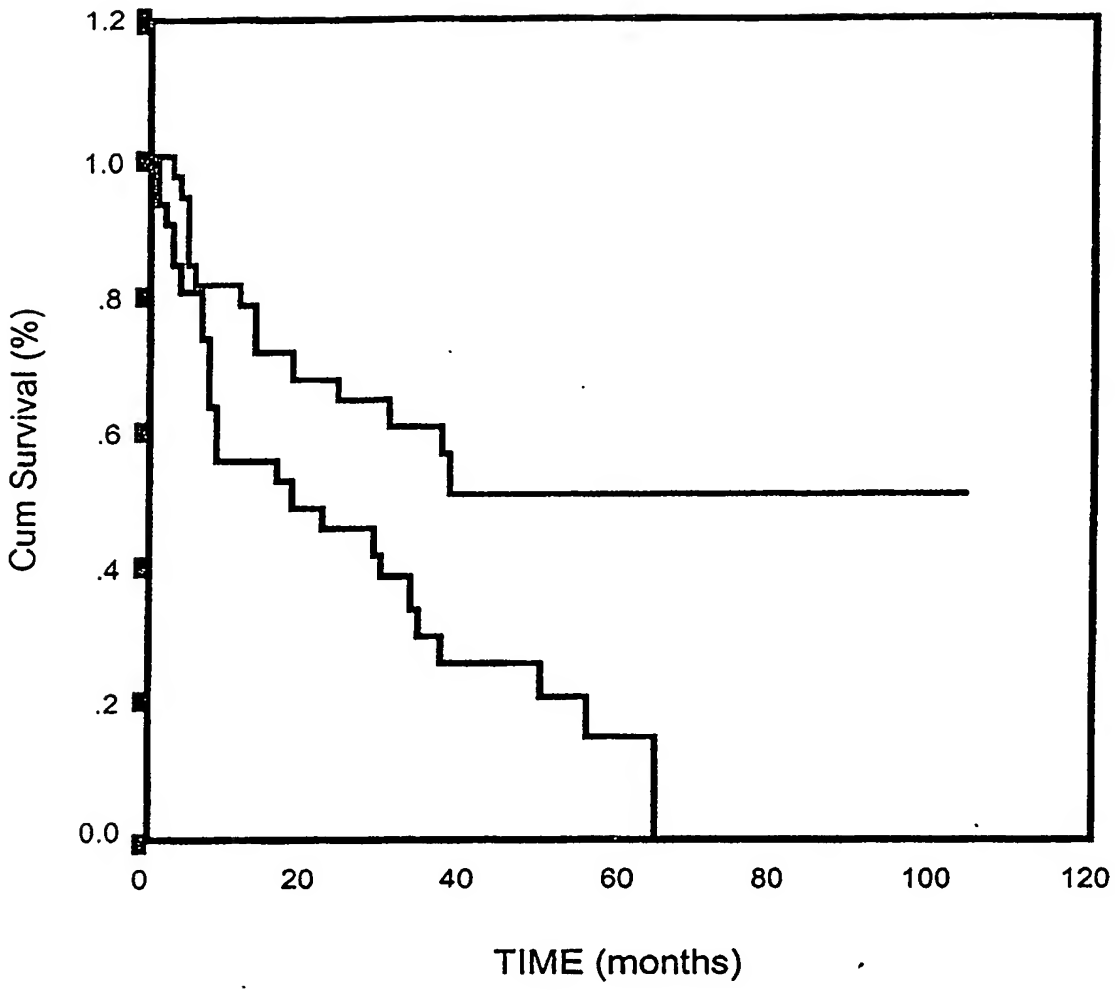

1.2
1.0
.8
.6
.4
.2
0.0
Cum Survival (%)
0
20
40
60
80
100
120
TIME (months)
p33ING1
p33ING1>20
censored
p33ING1<20
censored

Fig. 7C

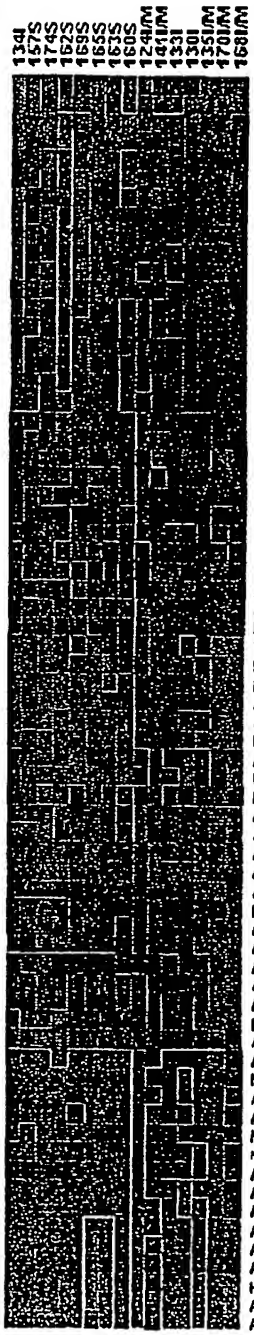

134
157S
174S
162S
169S
165S
163S
160S
124M
141M
133I
130I
135M
170M
168M
AA476543 similar to rat HREV107-like protein
AA455302 p33ING1
AA291556 Human ras inhibitor mRNA 3' end
AA485867 similar to antigen WCL1 [Bos taurus]
AA039231 calmodulin-related protein NB-1
AA598814 ATPase, Na+/K+ transporting, beta 1 polypeptide
AA129135 Human placental equilibrative nucleoside transporter 1
AA463480 similar to protein-tyrosine phosphatase
AA280748 p84 protein
AA487895 similar to similar to ankyrin
AA284277 similar to microtubule-vesicle linker CLIP-170
T55931 Human leukemia virus receptor 2 (GLVR2)
AA676590 Homo sapiens TTAGGG repeat binding factor 2 (hTRF2) mRNA c
AA699732 Homo sapiens hepatocyte nuclear factor 6 (HNF-6) mRNA part
T90767 Homo sapiens cyclin T mRNA complete cds
AA664077 Human fetus brain mRNA for vacuolar ATPase, complete cds
AA676804 CCAAT/enhancer binding protein (C/EBP), gamma
T94293 Human calcium-dependent group X phospholipase A2
AA046424 similar to ZAP128 protein
T50661 similar to HLA class I
AA102130 ESTs, Highly similar to actin-binding protein [M.musculus]
AA659341 p23 protein
AA284292 Glycoprotein-4-beta-galactosyltransferase 2
HB2532 similar to FRG1 gene product [H.sapiens]
T72422 Human tryptophan oxygenase (TDO) mRNA complete cds
AA974848 Homo sapiens myotonic dystrophy associated protein kinase
W44558 similar to dadrin coat assembly protein AP19
AA521469 Human putative G-protein (GP-1)
AA291959 dynamitin protein
N33331 Human peroxisome proliferator activated receptor PPAR,
T65864 Acyloxyacyl hydrolase (neutrophil)
H96554 Ceruloplasmin (ferroxidase)
N51705 similar to myosin-IXb [H.sapiens]
AA410604 Human CDC16Hs mRNA complete cds
AA009609 Fibroblast growth factor 7 (keratinocyte growth factor)
H68885 Homo sapiens IPL (IPL) mRNA complete cds
AA504266 H.sapiens mRNA for nuclear protein SA-2
W72051 Homo sapiens mRNA for fatty acid binding protein, complete cd
N45141 adenylate cyclase type II
AA149172 similar to 3-oxoacyl-[acyl-carrier protein
AA118524 Homo sapiens phospholipase D2 (PLD2) mRNA complete cds
AA609598 ESTs, Highly similar to 14-3-3 zeta/delta protein
AA157129 similar to Oxalyl Coa Decarboxylase
AA419693 Homo sapiens spleen mitotic checkpoint BUB3 (BUB3) mRNA
N73101 Diastrophic dysplasia (sulfate transporter)
AA504559 Cellular adhesion regulatory molecule [human, mRNA, 429 nt]
AA598629 large proline-rich protein BAT3
AA630374 H.sapiens mRNA for protein-tyrosine-phosphatase
AA424695 Integrin alpha-3 subunit
AA406020 interferon induced 17KD protein
R15740 Homo sapiens chondroitin-6-sulfotransferase mRNA complete cd
AA404694 Human focal adhesion kinase (FAK) mRNA complete cds
AA455012 similar to HSP1-A
N90783 Homo sapiens purinergic receptor P2Y5 mRNA complete cds
AA017443 similar to ZYXIN [Gallus gallus]
AA053637 palmitoyl thioesterase protein
N99049 similar to M-phase phosphoprotein 8 [H.sapiens]
N53380 Protein kinase C
AA644211 Prostaglandin-endoperoxide synthase 2
AA453607 similar to calmodulin
AA156907 similar to zinc finger protein 91
AA502897 Nuclear factor of kappa light polypeptide gene enhancer
AA452804 similar to GC-rich DNA-binding factor
AA670422 ADP-ribosylation factor 3
H52673 Bak protein
AA625080 acrosyn-trypsin inhibitor II precursor
AA464557 26S proteasome regulatory subunit p31

Fig. 8